DOCUMENT RESUME

ED 262 796 IR 051 307

TITLE IFLA General Conference, 1985. Division of Bibliographic Control. Sections on Bibliography, Cataloging, and Classification. Papers.

INSTITUTION International Federation of Library Associations, The Hague (Netherlands).

PUB DATE Aug 85

NOTE 127p.; Papers presented at the International Federation of Library Associations (IFLA) General Conference (51st, Chicago, IL, August 18-24, 1985).

PUB TYPE Viewpoints (120) -- Reports - Descriptive (141) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC06 Plus Postage.

DESCRIPTORS Artificial Intelligence; *Cataloging; *Classification; Databases; Foreign Countries; Indexing; Information Science; International Organizations; Library Associations; *Library Automation; Library Cooperation; Library Services; *Online Systems; Program Descriptions; Subject Index Terms

IDENTIFIERS *Online Catalogs

ABSTRACT

Papers on cataloging and classification which were presented at the 1985 International Federation of Library Associations (IFLA) conference include: (1) "Online Cataloging at the Bibliotheque Nationale (Francoise Finelli and Serge Salomon, France); (2) "Development of National Press Repertoire in the Condition of Multinational State: Its Trends and Peculiarities--on the USSR's Example" (N. P. Igumnova, USSR); (3) "Implementing Changes in Cataloging Rules" (Michael Gorman, United States); (4) "Project ESSCAPE--Expert Systems for Simple Choice of Access Points for Entries: Applications of Artificial Intelligence in Cataloging" (R. Hjerppe, B. Olander, and K. Marklund, Sweden); (5) "Classification in the USA" (Robert P. Holley, United States); (6) "Dewey Decimal Classification Online Project: Integration of a Library Schedule and Index into the Subject Searching Capabilities of an Online Catalog" (Karen Markey and Anh Demeyer, United States); (7) "Subject Indexing in France: Tools and Projects" (Suzanne Jouguelet, France); and (8) "Classification in Online Systems: Research and the North American Perspective" (Nancy J. Williamson, Canada). (THC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

IFLA GENERAL CONFERENCE, CHICAGO

1985

Division: Bibliographic Control

Section:

74-BIBCO-1-E

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED262796

On-line cataloguing at the Bibliothèque nationale

BEST COPY AVAILABLE

by

Françoise FINELLI

Head of the Bibliographic and Technical Coordination Center

Serge SALOMON

Project Manager of the Bibliographic Data Base System

BIBLIOTHEQUE NATIONALE
Paris, France

KEYWORD(S):

Bibliographic system
On line cataloguing
Organization

IR051307

For internal use only: Si

To be brought to meeting no: 65

Remarks:

to be presented orally ☒

for distribution only ☐

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A.I. Van Wesemael

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abstract

The Bibliothèque nationale has undertaken a programme of automation, whose first objective is the creation of a bibliographic data base.

The main characteristics of the bibliographic system are the file structure, including both the authority and bibliographical files, and the data structure, based upon the links between records. It provides new solutions for cataloguing problems, as well as powerful search tools and the feasibility of various products.

The basic goal of the system has been to give the greatest importance to on-line cataloguing. Its use radically changes the way in which a cataloguer works, moreover, it improves his efficiency, as well as the quality of the final product. Implementation of the system allows for a much greater flexibility of organization. On the other hand, the system requires the development of a structure of coordination which supervises the use of the system in the cataloguing departments and guarantees the consistency of the data base.

The project began in April, 1984 ; the system will be put gradually into operation between the fall of 1985 and the spring of 1986, when it should be accessible to both french and foreign libraries.

On-line cataloguing at the Bibliothèque nationale

Introduction

The Bibliothèque nationale has since 1981, been committed to a extensive programme of modernization through the gradual automation of its main functions, bibliographic as well as technical, financial and administrative. As a result, the majority of the library's staff have had to change their working habits.

The main project is the constitution of a data base which includes all bibliographic entries for documents from the Dépôt légal, acquisitions, gifts or exchanges. The system is scheduled to begin operation in the fall of 1985. In its initial stage, it will manage only french and foreign monographs and serials. The base will later be expanded by including documents of other kinds (records, engravings, maps...) and by retrospective conversion of manual files. By mid-1987, the system should contain the entire catalogue of printed materials since 1970, that is to say over one million titles.

The computer system for handling this data base, is designed as a cataloguing and searching system completely on-line. The software to be used in this which is totally new, has been developed by GEAC, according to the specifications of the Bibliothèque nationale.

The present study will endeavour to explain :

- the main functions of the system, through a study of its bibliographic capabilities which are new compared to those of existing systems.
- the implementation of the system and the resulting modification of work habits and organisation.
- The current status of the project.

Obviousy, each initiative involving automation differs according to the objectives of the organization, as well as its size, its daily operation and its technical requirements.

The Bibliothèque nationale is no exception. Although the system outlined below is certainly not the only or the ultimate answer, it is, nonetheless, in our opinion, a valuable contribution and a progressive step in the evolution of bibliographic systems.

We also feel that certain organizational steps taken by the Bibliothèque nationale might be of interest to other institutions.

It is equally our hope that this paper, and the ensuing discussion, will assist us in distinguishing bibliographic solutions particular to the Bibliothèque nationale, and those which might be generalized to medium and large-sized libraries.

The conditions of the project

We would like to begin with a brief summary of the conditions and constraints of the project :

condition 1

The Dépôt légal is one of the main functions of the Bibliothèque nationale. Consequently, it receives all material published in France, and, in exchange, has to make the bibliographic process for these documents. In this way, the Bibliothèque nationale receives approximately 40,000 new books and periodical titles published domestically, as well as roughly 30,000 books of foreign origin, obtained either through acquisitions, gifts or exchanges.

Therefore, the computer system at the Bibliothèque nationale must be considered above all as a tool for the production of bibliographic information.

In this respect, the Bibliothèque nationale differs greatly from a large number of libraries, who, instead of creating their own catalogue entries rely mainly on those existing in the data bases of the main north american networks.

Condition 2

The Bibliothèque nationale is the National Bibliographic Agency, and manages a catalogue containing more than twelve million titles. In view of this fact, it requires an automated system capable of ensuring a data base of high bibliographic quality.

Condition 3

As the principal source of bibliographic information in France, the Bibliothèque nationale sought to have as soon as possible the tools to distribute this data to other libraries in France as well as abroad. This, however, was a secondary consideration, priority being given to the development of both the cataloguing and data base management aspects of the system.

Lastly, the project had to take into account that, as a result of earlier failures in automation, the Bibliothèque nationale was considerably behind other large libraries in this respect. In a period of a few years, a complete change from an almost completely manual system to one that is totally on-line would be necessary. Therefore, organizational and human considerations played a crucial role in the development of this project.

I The functions of the system

We will limit our description here to the bibliographic system itself, which represents a completely new and original development in satisfying the needs of the Bibliothèque nationale.

It is the first phase of a system which will also handle inquiries by the public and the processing of the Dépôt légal.

The existing GEAC modules for acquisitions and circulation will soon be installed at the B.N.

a) Data structure

We will first present the structure of the database, as its conception underlines the richness and the power of the system and guarantees the quality and the consistency of the data.

The data base is made up principally of two types of files :

- the authority files
- the bibliographic file.

Each record is identified by a single number managed by the system called "Record Control Number" (RCN).

The content of each record is structured and coded according to a MARC-type format. The B.N. has chosen INTERMARC for the input and the processing of data, while UNIMARC will be the exchange format for bibliographic records.

The authority files ensure the normalization of the main access points to the bibliographic records. Each authority record holds the heading (with the possibility of having several "parallel" headings depending on, for example, language or usage current or scientific), as well as rejected forms and related forms.

Four authority files exist : authors personal names and corporate bodies, uniform titles and subjects.

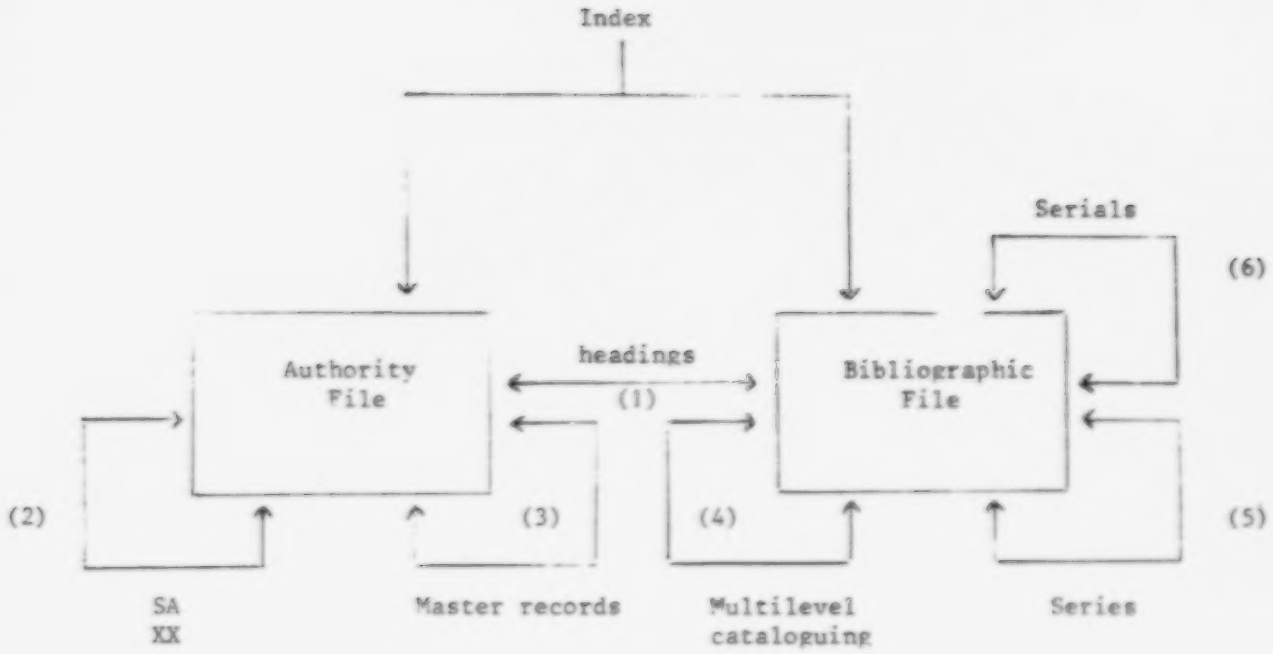
The bibliographic file holds records in a format according to the type of the document. At the outset, the bibliographic data base will contain only monographs and serials. Series are considered as serials in the bibliographic files (see appendix I).

Records are linked either through a link tag holding the RCN of the record to which the link is made when the link is from one record to one other record (eg See also reference) or through a "link index" when the link is made between one record and several others (eg linking one series to the N volumes of that series).

The database structure (see figure 1)

Figure 1

The logical structure of the database



BEST COPY AVAILABLE

The main links between records are the following :

1) Between an authority record and any bibliographic record which refers to it by a heading tag.

Within the authority file

2) Between two authority records when the heading for one appears in the other as a see also reference (SA) or tracing (XX)

3) Inside a given authority file, between an element common to several headings for which a master record has been established and its embedding in the heading of an authority record.

Within the bibliographic file

4) In the case of multi-level cataloguing between a lower level record and the record immediately above it.

5) For the cataloguing of volumes in a series, between the record of the volume and that the series.

6) Between the record of any given serial and that of another related serial (eg. previous title, subsequent title, edition, supplement).

These links are established through specific commands during cataloguing. They are all reciprocal and managed automatically by the system once a link is made in one way.

Furthermore, indexes allow the user to access directly into the authority and bibliographic files. In the case of information common to both files, the index is unique. In other words, an inquiry of the bibliographic file using a rejected form held only in the authority file, will nevertheless lead to the retained heading.

Both the indexes and the links are managed on-line, just after the "saving" of the record.

Only indexes by key-words and certain kinds of reciprocal links are created in batch.

b) <u>The cataloguing function</u>

The cataloguing function allows for the bibliographic handling of the following problems :

- the automatic control of the headings of a bibliographic record against the authority files. This feature is compulsory at the Bibliothèque nationale.
- the management of See also references and tracings of the authority records by a control against the authority files similar to that of the headings of a bibliographic record.
- the management of series through the creation of a link between the record of a volume in a series and the record of the series with the capacity to display in numerical sequence all the volumes in the series.
- multi-level cataloguing and analytical entries using a tree structure which is automatically processed by the system. Each record is treated, at any given level, as though it were a simple monograph (authority checks, indexes, links, etc...) but the display of such a record would necessitate the reconstruction of all the bibliographic information contained therein, beginning at the top of the tree.

- the management of links between serials (previous, title, subsequent title, supplements, editions).

c) The inquiry function

For librarians, inquiries are made through a series of commands. These inquiries make possible :

- access by different indexes, either to the authority files or to the bibliographic file. The indexes are the following : authors, subjects, titles (including series titles), ISBN, ISSN, CDU, entry number, call number...
- access through record links, for example, between one authority record and all related bibliographic records, or between one authority record and another to which it is linked by a See also reference.

These two kinds of accessing methods make catalogue scanning more dynamic.

Search keys can be truncated. Inquiries using keywords, and boolean searches, which combines several indexes by operators such as and/or/except are also possible.

Lastly, a record might be displayed in different formats :
MARC, including ISDS for serials
ISBD or edited form for authority records
Public, full or brief.

d) Products

Products are generated in batch.
In its initial phase, the system allows the production of :
- exchange tapes for bibliographic records of monographs in UNIMARC, authority records in INTERMARC, and serials in ISDS.
- COM tapes for authority file microfiches.
- tapes for photocomposing of current bibliographies, most notably the Bibliographie de la France.

II The use of the system and its consequences

a) On-line cataloguing

The main goal has been to give priority to on-line cataloguing capability to the cataloguer through the use of a terminal. So this end, the system features helpful cataloguing tools :

- terminal input of the record, as well as on-screen word processing. The input is made in full screen edit mode with possible scrolling of the screen. The system anticipates the cataloguer through the display of mandatory tags or sub-fields codes and values by default. The "local" mode is retained until the SEND key is pushed. This feature permits unhindered work at the terminal, until the operator is satisfied with the entered data.

- control of headings vis-à-vis authority files. The cataloguer may, if he so chooses, enter a truncated form of the heading. The system searches in the authority file for all forms which either begin with this string or are close to it. The cataloguer merely selects the heading he wishes to retain from the index displayed. If the cataloguer input a rejected form, the system will propose the accepted heading.
  Moreover, if a new heading is desired which does not have any reference in the authority file, the system will automatically generate a temporary authority record.

- controls of record format, content and consistency.

- creation of links using the automatic transfer of information into the record in process (for example, if a link is made with the record of a series, the system will transfer the key-title and the ISSN of the series).

- on-line index update.

Cataloguing and search procedures are totally integrated. At any time during the creation or update of a record, a search may be undertaken without risking the loss of the information already input.

For example, while entering a bibliographic record, a cataloguer may begin to key in the subject headings, break off in order to search the subject authority file, find the right authority heading but wish to its usage by checking its use in the bibliographic records, and, upon deciding to retain that heading, return to the record in process, link it to the retained authority record through the use of a link command and then continue his input.

b) The value of the system in the cataloguer's work

The cataloguer will find that the on-line system profoundly changes his work :

- whereas the cataloguer once worked alone without any immediate response, there now exists a continuing dialogue between the user and the system all along the process of a record. The system responds to the input instantaneously, and the cataloguer according to the messages given by the system may check or modify the data keyed in.
  Not only are formal checks possible, but also suggestions concerning links to other records or the logical sequence of commands. The record is, therefore, the product of a collaboration between the user and the system.

- the cataloguer has access to all the files stored in the data base at his terminal, which allows him to carry out the greater portion of his bibliographic work in one place. This is in contrast to the previous method, which required the individual to be constantly on the move between different manual files. In addition, he has a number of highly effective research tools, such as the various indexes, keywords and boolean search.

In sum, the use of the on-line system alters the structure of the cataloguer's work, improves its quality and efficiency.

c) Consequences of the System for organization

Contrary to previous practice, every user has, as a result of the on-line bibliographic data base, immediate access to all the information recorded. It eliminates multiple local files, and the delays inherent in manually updating them.

As a result, the Library is given greater freedom in organization.

To illustrate the efficiency of the present system, we outline below two models of organization based on opposite premises, integration and separation, which are possible and profitable only because of the existence of the new bibliographic system.

- Until now, the various pieces of information comprising a bibliographic record were produced by different departments or groups of people, who, thereby, constituted several steps in the creation of the final record : descriptive part, corporate body headings, management of series, attribution of a call number, indexing according to subject, etc. The result was that, besides from complex circulation of documents and delays in the cataloguing process, cataloguers found themselves specialized in clearly defined duties. From now, the distribution of the bibliographic information allows integration of these different tasks. We feel that the individual responsible for descriptive cataloguing will be able to assign and even create headings which would then be approved by those responsible of the management of the authority files. In the middle term, we believe that one person will be responsible for descriptive and analytical cataloguing and will be able to carry out the entire bibliographic analysis of a document. This will result not only in minimal circulation of documents and a greater speed in their processing, but also in the expansion and enrichment of the cataloguer's work.

- In some instances, maintaining a distinction between departments which share some functions in necessary. Such is the case with entry services, Dépôt légal and acquisitions, which, apart from their specific functions, have an indirect bibliographic role in the need to provide summary descriptions.

The proposed system allows the cataloguing department to use the minimal record made upon receipt of the document, which is stored in the data base and therefore available to any user. It can be reexamined and upgraded by the processing department, and then returned to the data base as a replacement for the first entry. In this case, the departments remain separate for functional purposes, while taking benefit, thanks to the system outlined here, from the accrued output of work carried out by all sections of the library staff.

A common data base and its availability at any terminal requires a greater inter-departmental coordination regarding methods and standards ; it is imperative to scrutinize the quality of every aspect of production. To this end, the Bibliothèque nationale has established a Bibliographic and Technical Coordination Center.

This Center is responsible for the bibliographic and technical consistency of the data base. It combines the functions of handling authority files, as well as standardization, documentation and training, with the administration of the data base. The data base administrator's functions include : education of the departments in the use of the system, investigation of functional and worker/work environment difficulties, proposal of changes and guarantee of the consistency of the data base.

The center has the means to combine bibliographic tasks (standardization, authority files) with technical work (supervision of the data base), to monitor the use of the entire system in each department and supply the necessary assistance (documentation, training).

The main object of this model is the effective creation of a useful link between the center and the different services involved in the process of a document. The support of processing departements in their daily use of the system is the prime function of the center. A core group of people will work full time at the center, supplemented by personnal who will divide their time between the center and a cataloguing service. Representatives responsible for surveying authority files and computer advisors will be chosen in the cataloguing departments, who will be under the technical authority of the Coordination center, despite their administrative link to the cataloguing service. They will be asked to follow the implementation of the system both in its bibliographic and technical content. These individuals will thereby assure a line of communication with the Coordination centre and will assist their colleagues in cataloguing (creation of headings, training, validation of records, etc.)

See figure 2.

III The present status of the project

a) Methodology

The BN-GEAC project commenced in April 1984.

At the Bibliothèque nationale, a project team comprised of a computer technician and a librarian, that is to say us, was chosen to define the requirements of the system and to follow its implementation by GEAC staff, located in Bristol. We worked very closely with the project manager at GEAC during meetings in Bristol which were held every six to eight weeks. The Bibliothèque nationale made important contributions to the project in terms of human resources from beginning to end. We reported regularly to a Bibliothèque nationale pilot committee comprised of the department heads most affected by the system. This committee was responsible for approving every important decision made in the course of the project regarding the practical operation of the system, as well as the timetable for its installation.

Supplemental to these committees, a BN-GEAC steering committee, including the General Administrator of the Bibliothèque nationale and the Vice-President of GEAC Canada meets roughly every four months.

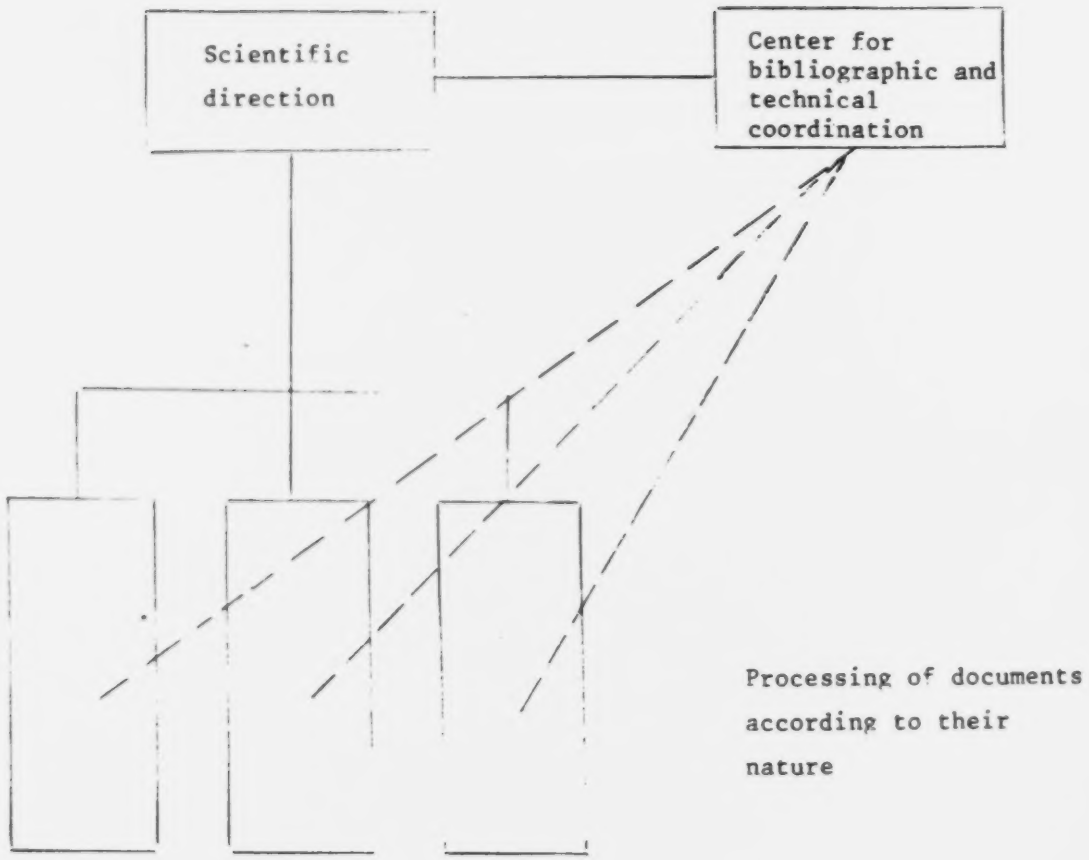
b) The temporary use of existing software

While awaiting the arrival of the new system, the Bibliothèque nationale underwent several experiments with existing GEAC software :

To begin with, the file of the Bibliographie de la France 1975-1984 including a subject index beginning in 1980 is now on-line. As well the authority files were progressively entered into the system in order to allow not only user inquiry but also the creation and the updating of the records.

The authority records for corporate bodies and personal names are the product of the input of the worksheets made by the relevent sections combined with headings and cross-references extracted from the bibliographic records of the Bibliographie de la France. This technique has been used in the compilation of over 30 000 records of corporate bodies and 120 000 records of personnal names. In addition, the authority files for subjects including headings and subdivisions (see Appendix I) was loaded after the conversion of an existing file of 65 000 records.

Figure 2


Scientific direction
Center for bibliographic and technical coordination
Processing of documents according to their nature

The results of this experience with the authority files have been very positive : cataloguers learned how to deal with the format and use of a computer system. It also afforded the opportunity to define the methodology for working on authority files through practical application. Above all, it allowed for an upgrading of a large number of records and, consequently, the improvement of the quality of the authority files, which will constitute the foundation of the bibliographic data base.

c) The implementation of the system

At the time of writing, the new system is still undergoing tests. The core of the system exists, but some details are still to be completely worked out.

The acceptance of the system will start by September 15. If all goes well, the existing files will be loaded into the data base. At the outset, it will contain 350 000 bibliographic records and 220 000 authority records.

The system will be gradually introduced into the departments : From the date of the end of the tests (in theory, October 15, 1985), or Day Zero, the cataloguing of the following documents will be automated :

Day Zero : Official publications, foreign monographs
Day Zero + 2 months : French monographs
Day Zero + 4 months : French and foreign serials

Final receipt of the software will coincide with the moving of the processing services into the new building of the Bibliothèque nationale, where roughly a hundred terminals will be available for use. By end of the year, over 150 individuals will use terminals in their work.

Conclusion : **Prospects for application of the project outside the Bibliotheque nationale**

Beginning in early 1986, and according to the expansion of the system, the Bibliotheque nationale hopes to offer the following services :

- various batch products : current bibliographies, microfiches of the authority files, UNIMARC exchange tapes.
- on-line inquiries of the data base by other libraries using any standard-type terminal through leased lines or by TRANSPAC.

The Bibliothèque nationale system should rapidly become accessible to other libraries, either in order to download files on their own system, or add holdings to existing records or to create new ones as part of a union catalogue.

Connection of the system with those of other libraries in France or abroad is being studied, taking into account the application of standards for the exchange of bibliographic information and the interconnection of systems. Such an interlibrary effort is currently being undertaken by the Bibliothèque nationale and the British Library and should be in operation by the end of 1986.

Thus, the predetermined objectives of the project are becoming reality according to the preset schedule. Nonetheless, there are obstacles to be overcome. For example, the period from the fall of 1985 to early 1986 will be extremely difficult. However, the Bibliothèque nationale has full confidence in the success of the project, which will put an end to the wait by French libraries and give new life to our collaboration with major foreign institutions.

APPENDIX I

In our elaboration of the cataloguing system, we made certain choices which may provoke discussion, namely :

The authority file for authors, personal names.
The authority file for subject headings.
The processing of serials.

Is it worthwhile to maintain an authority file for personal name authors ?

The major part of our authority file was extracted from the headings of bibliographic records in the Bibliographie de la France, 1975-1984.

We can compile the following statistics :

From a file of 210 000 bibliographic records, 115 000 authority records were generated : therefore one authority record was created for every two bibliographic records. 3 500 authority records are linked to more than 10 bibliographic records, a figure representing only 3 % of the cases.

In view of the small percentage of personal name authority headings used, one could well be expected to question the necessity and the benefits of maintaining an authority file.

An initial response would be that the composition of the authority file makes possible a flexibility unattainable if the information were stored only at the level of the bibliographic record.

- it is possible to record several headings, such as the Latin form "Livius, Titus" and the current French one "Tite Live", which allows user libraries to select the term best suited to their own file.

- it is possible to link several pseudonyms used by an author in creating an authority record for each and linking them by a See also reference.

A secondary response is the need for clear bibliographic information.

In working with the authority file made from the extracted headings of bibliographic records, we found that there was no distinction made between homonyms. For example, one index entry, "Adam, Paul" was linked to seven bibliographic records, but referred to four different authors. As a result, we created four authority records whose headings were qualified by dates or professions. An authority file is the only way to systematically account for and separate homonyms.

In terms of our goal of greater ease for the user, it may seem tedious to be obliged to verify every heading of a bibliographic record in the authority file. The system, as conceved, should not hinder cataloguing in process. Indeed, the cataloguer may enter the name which appears on the work. The system will automatically scan for this form in the authority file and then display an index including the corresponding heading if it already exists, the reference to the heading if the input form was a rejected one, or, if no entry is exactly matching, an alphabetical index of the headings most closely resembling it.

The cataloguer may confirm the proposed choice in the first two instances. With respect to third, he may, without further searching, confirm the heading entered : the system will create a temporary authority record. The cataloguer should not have an increased workload because of this procedure.

The duty of the authority file supervisors is, as we have seen, mainly to differentiate between homonyms. While it is not possible to review each and every new heading, this may be carried out according to the number of times a heading recurs. If an authority record is connected to more than, for example, ten bibliographic records, it should be systematically checked in order to guarantee that this one heading is not, in fact, several distinct headings. This should ensure the bibliographic quality without slowing down the work in progress.

Towards a new management of the authority files for subject headings : from subject headings to key words

Since 1980, the Bibliothèque nationale has used a form of subject indexing based on the Répertoire des vedettes matières originated at Laval University, Québec, which is itself, in a large part, a translation of the Library of Congress subject headings. In the beginning, we composed an authority file generally basedon the subject headings used for the indexation of documents. Taking advantage of the changeover to the new system, we decided to modify the structure of the authority list : henceforth, in the authority, we will no longer oversee complete subject headings, but only the sub-elements, such as the main heading and the subdivisions. Only combinations requiring references or notes, or those in which the subdivisions are relevant only to the main heading to which it is linked will be kept as headings of authority records. It is our belief that this will facilitate the management of the subject authority file.

In order to handle the headings which correspond to a combination, where one of the elements, either the main heading or a subdivision is employed frequently in the file, we have decided to use master records. A master record is a specific authority record which may be linked to other master records or to authority records, but not to bibliographic records. It is a tool for the internal management of the authority file. When the heading of a master record or an authority record includes an element of a previously created master record, a link is established between the element in the relevant heading and the master. The link is made through the ACN (Authority Control Number) of the master. If a change is required in the heading of the master, for example a change of United Kingdom to Great Britain, one can easely locate all the related headings and modify them automatically.

The structure of the authority file has consequences for the subject indexing of the bibliographic records. Frequently, the subject headings made for a bibliographic record will not have an exact match in the authority file, although each of its elements, the main heading and the subdivisions, may have a direct match. The autority check will be carried out on the character string and the nature of the element (a given subdivision in the authority file corresponds to a heading with a particular tag in the authority file), but the overall syntax of the subject heading will not be checked. We feel that a partial solution as the subject index to the bibliographic records may be used as a model for the cataloguer. We are aware of the possible risk of errors. However, we do believe in the necessity of this type of indexing for a printed catalogue, although on-line inquiries will be made, in most cases, using key words, regardless of the order given by the cataloguer. Our intention was to

retain the ability to make correctly filed printed products in keeping with LC/Laval principles while moving towards key word indexing.

The management of series : bibliographic file or authority file ?

We have treated the series file as a bibliographic file, as they are regarded as a subset of serials with their own particularities. Instead of creating an authority file for series titles, we included the main functions for the management of authorities in the bibliographic record. For each document in a series, the format contains two fields, one for the ISBD series statement, the other for the link with the series bibliographic record. When the link between the record of the volume and that of the series is established, the descriptive field corresponding to the series statement is checked against existing title fields in the series record (key title, proper title, parallel titles, variant titles). If there is no match, the title as listed in the series statement of the volume is added as a variant in the record of the series. Verification of the title in the series statement among the titles in the series record does not lead to its replacement by the "heading" title which should be the usual authority check, but to its addition as a variant when neccessary. In the terms of the authority record, the variant title is equivalent to a see reference, while linking with other series (series/subseries, preceeding or following title, etc.) are the same as see also references. Therefore, the bibliographic record of the series includes all the information that should be found in an authority record.

APPENDIX II

Figures relevant to the data base of the Bibliothèque nationale :

- Early 1986

400 000	bibliographic records
250 000	authority records
100	on-line cataloguing terminals
1	transaction every 1,5 secs.

- Late 1987

Over one million	bibliographic records
500 000	authority records
250	on-line cataloguing terminals
1,5	transactions per second.

N.P. Igumnova,
Deputy Director

Development of National Press Repertoire in the Condition of Multinational State: Its Trends and Peculiarities /on the USSR's Example/

The retrospective national bibliography developing in the conditions of the Soviet multinational state has its peculiarities. Each member-republic of the Soviet state compiles its national press repertoire which is an integral part of the entire state system of the Soviet bibliography.

The national press repertoire reflects the documentary stream of the Soviet multinational state most adequately and is oriented towards satisfaction of the most universal information needs of society. Taking special place in the country's bibliographical production, the national repertoire is the base for preparing sub-scientific, recommended, branch, production, local lore and other bibliographies.

The national press repertoire is a significant contribution to the development of the Soviet multinational culture and to the building of culture in every republic; it furthers a more extensive utilization of the mankind's spiritual wealth as well as of the achievements of science, culture and production fixed in printed matters.

Before we consider some basic trends in the development of this kind of bibliography, let us dwell on the terminology used in our country. The term of the "national bibliography" is conditionally used mainly for the communication at the

international level, since the registration bibliography has a clear multinational character. Yet, a wider spread term is that of the "state bibliography" which means both the entire state bibliographical registration of published works and the national bibliography of each national republic. Rather frequently used is also the term of the "national bibliographic repertoire" which covers particular retrospective bibliographic indices and their aggregations containing full lists of printed works for a certain chronological period and of hand-written books since the emergence of a written language in the state. The concept of the "repertoire" presupposes the exhaustive plentitude of materials within certain boundaries. The inclusion of the hand-written books into the national press repertoire is substantiated by the content and design integrity of hand-written and printed books and by their social significance as objects and primary sources for studying material and spiritual culture of the Soviet nations.

Synonymous with the concept of the "national repertoire" is the concept of the "general retrospective bibliography" which also denominates the compiling of retrospective sources of bibliographic information embracing the state's total press production.

All these terms--"national", "state", "general retrospective bibliography", and "national repertoire"--actually, have the same content. The first two are most often used with regard to the current bibliography, whilst the two others--with regard to the retrospective bibliography. Since the term

of the "national press repertoire" is widely used in the world to denominate the national bibliography of a retrospective character, it will be predominantly used in this report, too.

Since the very first years of its existence, the Soviet state has been paying great attention to the elaboration of the national press repertoire.

A special decree of 1920 included measures to develop state bibliography, introduced an obligatory current registration of the entire press production and ensured its permanent archives storage by the All-Union Book Chamber and the book chambers of the republics. There was a decision adopted simultaneously about a centralized cataloguing of the "obligatory copy" printed works. There was a ramified system of current state bibliographic publications created in the country, differentiated as to the content, typology, department, territory and other indications. The All-Union Book Chamber and the book chambers of the republics were the main centers of the current national bibliography.

Yet, there remained a serious problem of the retrospective registration and of the introduction of the national repertoire of all the books, hand-written materials and documents which appeared in the territory of our multinational state in the course of Soviet nations' historical and cultural development.

In 1965, some very important documents were elaborated in our country which allowed to organize a retrospective bibliography system in our country: Directives on Retrospective General Bibliography of Printed Works Released by National

Autonomous Republics of the USSR in the Soviet Years, and Directives on Retrospective General Bibliography of Printed Works Released in National and Autonomous Republics in the Pre-revolutionary Period. The purpose of those documents was to orderly organize the retrospective bibliography, to ensure a full registration of editions reflecting the achievements of national culture, and, therefore, effectively help in developing science and culture in the USSR, in finding out necessary sources of information. From the very beginning the leading role in the creation of the retrospective national repertoire was played by the biggest scientific libraries which had the most complete collections of printed matters, developed reference and bibliography apparatus, skilled personnel, and rich experience of bibliographic work.

It is historically established that the retrospective national bibliography, at the entire state level, falls under the competence of such libraries as the V.I. Lenin State Library of the USSR, the Library of the USSR's Academy of Sciences, in coordination and cooperation with other biggest libraries of the country, whilst, at the level of national republics, it is done by state republican libraries which have the richest book collections of the past periods in the country. The reference and bibliography apparatus of these biggest libraries, scientific descriptions of these collections, as well as of private and other particular collections serve as an important source for the creation of the national repertoire.

As to the amount of the registrated printed production, the

retrospective indices of a repertoire character are especially fundamental and chronologically deep.

The national press repertoires of the republics include editions which reflect information about particular republics. They include translations of works by national authors into the languages of other nationalities of the USSR /or of the world/, literature about particular republics and their people published in other national republics and abroad and, completely or in part, connected with the life of this or that republic. These materials fall under the category of the "exteriorica" and include editions about prerevolutionary Russia, the USSR and its republics; reviews of Soviet scientific editions; translations into foreign languages, original works in foreign languages, etc. The exteriorica published outside the republic but linked up with it by the language, content and some other characteristics is often considered as a part of the national bibliography.

In the conditions of our multinational state it is especially important to observe the same principles and a common methodological approach while adopting decisions about the methods of compiling and organizing national press repertoires in the republics.

Basic indications of national repertoire bibliography in our country are: state-and-territorial, including the studies about the republic, and national. The majority of retrospective indices, with the inclusion of the prerevolutionary printed

production of the national republics, use either the language principle or the language principle in combination with the principle of the author's national background, and others. Thus some national retrospective handbooks are based on this set of characteristic features: territory, language, nationality and content. /Press of the Lithuanian SSR, 1940-1965, Vol. 1-2 /Vilnius, 1962-1980/, Book in Soviet Moldavia, 1924-1964 /Kishinev, 1974-1976/, Books of Soviet Estonia, 1940-1970 /Tallinr 1956-1978/ and others/. Such indices as Book in Soviet Armenia, 1920-1930 /Yerevan, 1978, Vol. I/, Lettish Periodics, 1768-1919 /Riga, 1961-1977, Vol 1-3/ include press production, i.e., everything released in all the languages in the territory of these republics and also reflect the works in the language of the indigenous population published in the territory of the USSR.

The majority of retrospective bibliographic indices gives bibliographical descriptions in the state language of the republic, with a necessary indication of the language in which the work was printed, and the translation of its title into Russian.

The objects for a retrospective registration of printed matters are books, booklets, journals, newspapers, continued editions/works and others/, musical compositions, fine-arts works of a multiple circulation, maps and atlases, as well as some materials of not a wide spread.

A peculiar feature of this kind of bibliography is the character of bibliography-making. National bibliography follows international recommendations, the State Standards and Common

Rules to Describe Printed Works. Due to a colossal volume of bibliographied materials in our country, a minimum list of requirements has been worked out: a strictly limited number of elements of bibliographical entries, simplified structure, etc. which, in particular, facilitates introduction of automation. This approach does not rule out any profound processing of materials concerning such objects of bibliography-making as old-print, valuable and rare editions, including illegal publications of the prerevolutionary period in Russia, manuscripts and ancient hand-written books.

It is expedient to consider also some basic functions of the national press repertoire. Repertoire bibliography fulfills, primarily, the following two functions: that of communication and that of a quest for books. It has also some other functions:

- cultural and historical which characterises the cultural development of nationalities in the Soviet Union;

- statistical which ensures the taking stock of and registration of printed and hand-written documents, the revealing of all the documentary resources of the country;

- address which allows to locate the bibliographied materials.

The national press repertoire of our country exists, mainly, in the form of summary catalogues which allows both to know where a particular edition is located and to open up and propagate collections of the biggest state book depositories.

While considering basic characteristic features of the national press repertoire, it is worth mentioning that as to the

chronological boundaries of the inclusion of material, national retrospective indices take into account the printed production and the manuscripts of the entire period of the existence of the written language and of particular periods of the state's history. As to the wide scope of the content of the bibliographied material, these are universal text-books, without any limitations of content or theme. Being prepared alongside, are some fundamental problem-and-theme-oriented bibliographic editions, indices on one branch of knowledge, particular kinds and categories of publications serving as supplements to the system of national bibliography indices. As to the language of the objects to be bibliographied, there are text-books reflecting the works published in the given territory, in all the languages or only in one. As to the place of publication, the retrospective bibliography reflects the local press works of a certain region of the country and of the country's entire printed production. International bibliographies of a repertoire type /incunabula indices, indices of old-print Slav and of handwritten books/ should be mentioned as well. Moreover, there are indices of printed production and of particular publishing- and printing-houses released, too. As to the authors, there are indices published which take into account the works by all the authors.

The most widespread form of indices of the repertoire type is printed bibliographical editions. Their basis is card catalogues and card indices. In future more attention will be paid to microforms and machine-read databases.

In our country national press repertoire of the republics

has, in fact, been created. The compiling of a complete repertoire of the Russian book still remains to be a complicated task. In 1947, a special document was elaborated which played an important role in the creation of the national press repertoire of the USSR. It was a Directive about Summary Catalogue of the Russian Book /1708-1947/ approved of by the Commitee on Cultural and Educative Institutions under the Council of Ministers of the RSFSR. The Directive determined the Summary Catalogue as the most topical part of the General Summary Catalogue of Russian Press which provided for registration of different kinds of printed works /books, periodics, cartographic, music, fine-arts, and other editions/. The Directives provided for the creation of the Summary Catalogue of the two periods--the prerevolutionary one and the Soviet one, beginning from 1917. The card base was created for the purpose by the All-Union Book Chamber /AUBC/, the V.I. Lenin State Library /LSL/, the M.E. Saltykov-Schedrin State Public Library /SPL/, and the Library of the Academy of Sciences /LAS/. The Summary Catalogue is arranged in an alphabetic order, according to the periods: 1801-1917, 1917-1927, 1927-1957. The Summary Catalogue of the Russian Book of the First Quarter of the XIX Century is being prepared for publication now.

The existing system of the retrospective state/national/ bibliography is based on the principles of coordination and cooperation of the biggest library-and-bibliographic institutions of the country: the AUBC, the LSL, the SPL and others,

the LSL fulfills the functions of the scientific-and-methodological and organizational center, works out recommendations on the methodology of compiling text-books, bibliographic description of the included editions, ensures standardization and unification of the main processes in the preparation of the retrospective national press repertoire. The RSFSR State Public Library of History /SPLH/ and the Scientific Library of Moscow State University /SL MSU/ are active participants of the system.

The LSL and the SPL are engaged in the bibliography-making of the books in Cyrillics and in recreating the repertoire of publications of the biggest printing centers of Moscow, Petersburg, Chernigov and some other cities, of the works of the first printers I. Fyodorov and P.T. Mstislavets, of the Belorussian printer Spiridon Sobol and others. A special trend here is made up by the scientific descriptions of the Ukrainian, Belorussian and South-Slav editions preserved in these libraries. The LSL, the SPL, the LAS and other libraries see their main task in showing the succession of cultural ties of different nations and are preparing a starting ground for the creation of the code of national press repertoires of our fraternal republics and of some countries of the socialist community. With this aim in view, the LSL, for example, interacts with all the other national republics and with the following socialist countries: the CSSR, the PPR, the PRB, and others.

In the USSR a systematic work is being carried out aimed at creating a retrospective code of periodical and continued editions. Now we already have a system of interconnected funda-

mental bibliographic editions which, in an almost uninterrupted chronological succession, reflect the Russian periodical press-- from the very first newspaper /1702/ up to the newest cumulative indices of newspapers published by the All-Union Book Chamber.

The retrospective stock-taking of periodicals of the prerevolutionary period is also done in the form of summary catalogues or the catalogues reflecting the corresponding repertoires of particular depositories. The result of the LSL's 25 years of activities is a fundamental work in 4 volumes which represents about 10 thousand Russian periodicals /1901-1916/. In the bibliography-making of the prerevolutionary periodicals, the LSL adheres to a definite topical orientation. Indices of the Russian Trade-Union Periodicals, the Russian Military Periodicals, the Russian Satirical Periodicals and an Alphabetic Index might be cited as examples. The LSL depositories are also being opened up. Thus there has been an index published-- Russian Prerevolutionary Newspapers in the Depositories of the V.I. Lenin State Library /1702-1916/, and some others.

On the basis of cooperation of the biggest library-and-bibliographic institutions of our country /the AUBC, the LSL, the SPL/ -- and with the participation of some others-- along with the working out of the national repertoire of the Soviet periodical press, an index Periodical Press in the USSR /1917-1949/ was released; it has many volumes and includes over 25 thousand titles of journals, works, and bulletins in the languages of the Soviet nationalities and in foreign languages.

In 1970 a reference book Newspapers in the USSR came off

the press; it carried information about the newspapers printed in the territory of the USSR in 1917-1960.

In the LSL set apart as particular depositories are such kinds of press works as cartographic materials, music editions, fine-arts materials and others. Specific character of such publications demanded from the LSL to elaborate special methods of their registration and bibliographic processing /descriptions, annotations, systematization, system of "keys" differentiated with regard to each kind of editions/.

In conclusion, it should be said that the peculiarities of the retrospective state bibliography are determined by the guiding principles of development of the Soviet multinational state fixed in the Constitution of the USSR; by the international character of this state, sovereignty of every national republic and the common character of interests of the Soviet people, by the interaction and mutual enrichment of national cultures in the USSR.

The LSL actively cooperates in the elaboration of some theoretical propositions with regard to the retrospective national bibliography and in the work at some concrete bibliographic text-books, at an international level.

Within the frameworks of the international comparative research National Library in State and International Information Systems /NL in NATIS/ carried out in 1981-1985 by the national libraries of the socialist countries /the CSSR, the DRV, the GDR, the HPR, the MPR, Republic of Cuba, and PPR, the PRB, and the USSR/ some basic trends in the activities of the NL's

of the socialist countries were determined--concerning the writing of text-books about the retrospective national bibliography /RNB/, its tasks and principles of organization; there was a draft directive on the Role of National Library in Establishing the Entire-State System of Retrospective Bibliographic Information prepared and approved of by the NL experts from the member-countries.

In the last decade--due to the elaboration of national and international information systems--it became necessary to determine the development of and to improve the retrospective national bibliography. These issues were included into the IFLA long-term programme, the LSL prepared a Summary Table of Publications of Retrospective National Bibliography of 64 countries of the World which gave an all-round analysis of 342 text-books reflecting different kinds of documents. Actually, the Tables serve as the basis for studying the state of affairs and detecting basic trends in the RNB's development.

The LSL is carrying a lot of work in cooperation with other national libraries of the world to create sources of bibliographic information. Thus the LSL, together with the NL's of Bulgaria, Czechoslovakia, the GDR, Hungary, Poland, with the NL of the republics of Yugoslavia /about 60 all in all/, is preparing a Summary Catalogue of Old-Print Editions of Cyrillic and Glagolitic Types. The work began in 1976. The first volume of the Catalogue is being prepared for publication.

The V.I. Lenin State Library, together with some libraries abroad, is studying the existing editions of national biblio-

graphy. In elaborating the theme of the NL in NATIS, the socialist countries' bibliographic production came under analysis. The LSL prepared an annotated index State /National/ Bibliographic Indices of Socialist Countries which is the source-study basis for this research. The NL's of Bulgaria, Cuba, Czechoslovakia, the GDR, Hungary, Poland, the USSR and Yugoslavia as well as the IFLA Section of Bibliography took part in the work at the Indices. Their purpose is to widely inform the library public, bibliographers, the faculty and students of the institutes of culture and of other higher educational establishments about the system of national bibliography editions existing in the socialist countries.

IFLA GENERAL CONFERENCE, CHICAGO

Division: Bibliographic Control

Section: Cataloguing with University Libraries

1985

132-UN-8/CAT-3-E

Implementing Changes in Cataloguing Rules

by

Michael GORMAN

Professor of Library Administration & Director,
General Services Department
University Library

UNIVERSITY OF ILLINOIS
Champaign-Urbana, U.S.A.

KEYWORD(S):

Cataloguing rules
AACR

For internal use only: 51

To be brought to meeting no:
170

Remarks:

to be presented orally ☒

IMPLEMENTING CHANGES IN CATALOGUING RULES

Cataloguing rules change. They change because the institutions and the agencies which are responsible for them are, sooner or later, forced to respond to changes in the bibliographic universe and to developments in cataloguing theory. Those changes and developments are, by and large, neither whimsical nor contextless. To take a very simple and practical example, the advent of the so-called non-print materials, the spread in publication and use of those materials, and their consequent acquisition in large numbers by libraries forced a re-appraisal of the book-oriented cataloguing codes of the past. In the realm of theory, the Anglo-American bibliographic control universe was profoundly affected by, successively, the ideas and writings of Seymour Lubetzky[1], the Paris Principles[2], and the International Standard Bibliographic Description (ISBD)[3]. The same forces and the writings of Domanovsky[4] and Verona[5] have influenced cataloguing rules and practice throughout the world. This internationalizing of cataloguing theory, and the universality of changes in the bibliographic universe, while being undoubtedly progressive and beneficial, has led to a level of conflict beyond that of the national debate which revolves around costs and

benefits. That other level is between those who take a local view of cataloguing problems and, hence, tend to view internationally motivated change as questionable, at best, and those who take the concept of Universal Bibliographic Control[6] seriously and, hence, tend to view all steps which harmonize cataloguing practice as being inherently desirable.

At all levels, there is a perceived clash between those who tend to view library problems in economic and managerial terms and those who believe that bibliographic systems have to change, that that change may be, periodically, seismic in nature, and that the cost of change is less important than the benefit. The caricature is of the library administrator who knows the price of everything and the value of nothing and the cataloguing expert who knows the value of everything and the price of nothing. As someone who was the secretary of the original ISBD working party, is an editor of AACR 2, and is a senior administrator of a major university library, I tend to reject the caricature. There is, in fact, a spectrum of opinion. Most persons interested in these matters take a position between those of the utterly cynical administrator and of the starry-eyed bibliographic dreamer. In these matters I am a Utilitarian. What tends toward the greatest happiness of the greatest number has always seemed to me to be desirable in bibliographic control as well as in other areas of life.

Before going on to examine the question of the implementation of AACR 2, I wish to refer briefly to the alternative to change in

cataloguing rules. That alternative is, of course, stagnation; and that stagnation rapidly leads to the death of co-operation. As cataloguing rules grow older and more out-of-date, they turn inward, and the institutions which use them and which elaborate upon them to cope with new cases become more and more peculiar in their cataloguing practice. Bibliographic history abounds with institutions which have built massive edifices of local practice on the slender foundations of out-moded cataloguing rules and have, therefore, been unable to co-operate in shared cataloguing because of the singularity of the records which they create.

The first skirmishes of the War of AACR 2 were fought long before that code existed and, in fact, long before it was even thought of. The Anglo-American cataloging rules of 1967[7] (the first truly international code for the English-speaking world since 1908) were to be the product of the Lubetzky revolution. Reactionary elements prevented that from occurring. Not only were certain outmoded practices held over because of pressures from large libraries, a circumstance which led the "Anglo" half of the partnership to rewrite the rules and thus, ironically, achieve a more truly Lubetzkian result, but also the code itself was never fully adopted by the American library community because of the now infamous policy of "superimposition." That policy, promulgated in a vain attempt to keep the Library of Congress card catalogue alive, stated, in essence, that no established heading would be changed and that only headings which were new to that vast catalogue wiould be framed in accordance with the new rules. The

obedient American library community followed suit. The result was bibliographic anarchy. The ostensible cataloguing rules were followed in the UK (in accordance with their own version) and the North Americans followed now one code and now another depending on the age of the heading which was under consideration and upon the guessing ability of a cataloguer in any library other than that of Congress.

Then came the ISBD. Certain elements in North America treated this innovation as a foreign plot[8]. Others sought to integrate it into their cataloguing practice[9]. The arguments were bitter and lengthy. The end result was that American cataloguing practice became schizophrenic. The headings came from many cataloguing eras and the description was as up-to-date as could be. Something had to give.

AACR 2[10] sought to reconcile all these irreconcilables. It was to be the tool by which superimposition could be ended, it was to introduce ISBD concepts into all areas of descriptive cataloguing, it was to be fair to all the various kinds of material collected by libraries, and it was to be the first truly international English language code. Amazingly enough, it did all those things. This did not guarantee a ready acceptance by the user community. I shall not here attempt to describe the conflict[11,12], but content myself with observing that the stage was set for the classic cost vs. benefit argument. Unfortunately, neither side had much in the way of objective data. As one person

armed with a guess (educated or otherwise) brawled with another with an opposing guess, heat was generated but little light was shed.

One important reason for the fundamentally futile nature of the War of AACR 2 was that there were other factors operating which became embroiled in the argument without ever being fully stated and, one suspects, without being fully understood by many. Briefly put, those factors were the slow death of the card catalogue in large libraries and the, in retrospect, untimely success of of the MARC idea. The former muddied the waters because changes in cataloguing rules became confused with the necessity to find alternative technologies for bibliographic control. The latter, with great irony, became a conservative force because of its great success. MARC, because of its historical position, became the embodiment of superimposition. It was the glittering coffin of a dead policy. Millions of Marc records contained flawed bibliographic data. The databases (OCLC and the like) built upon the success of MARC were faced with an urgent dilemma. Should they defend the past or fight for the future? If the latter, what about the cost?

What was the typical administrator (if there be any such) of a large library to do? He or she was, in all likelihood, faced with a moribund card catalogue, with a cataloguing department which, probably, either did not know what was going on or which knew but had not communicated that knowledge to the administration, and with no data (financial or otherwise) upon which to base an

intelligent decision. The temptation to say "nay" must have been near-irresistible. So, in the majority of cases, it proved to be. Questions such as the impossibility of continuing the policy of superimposition were ignored by many if not most. An Ils ne passeront pas mentality became the prevalent mode. The Library of Congress was persuaded to delay its adoption of AACR 2 by a year. Another year of "superimposed" records, of more pollution of the MARC databases, and of heads buried in the sand was, thereby, officially decreed. Eventually (in January 1981), AACR 2 was adopted. It has been in force for more than four years throughout the English-speaking world. The Library of Congress has closed its catalogue and has ended the the policy of superimposition. The American bibliographic networks are, by and large, thriving. Bibliographic contributions to, and use of, them is at an all-time high. Many large libraries have introduced online catalogues. In some cases, AACR 2 was the catalyst in the decision to change the form of catalogue. Western civilization seems to be in no worse, and no better, shape than it was before AACR 2. As Old Kaspar would have observed: 'Twas a famous victory.

What are the costs of implementing a new code such as AACR 2? They, of course, depend upon the circumstances of the library and of the bibliographic pond in which it swims. Broadly speaking, the costs are those of educating the cataloguers (and the users of the catalogue - staff and public) in the new rules, of providing connections between "new" and "old" entries, and of paying

(directly or indirectly) for networks and other national and local organizations to change their cataloguing practices. These are, by no means, sterile costs. By this I mean that the education of library staff and library users is never a wasted activity, that any reconsideration of established practice is healthy, and that all useful connections within the catalogue or between catalogues are worthwhile. None of this is to say that those costs are always acceptable or that the money to pay them is always available. The latter is a function of the fact that libraries rarely budget for the inevitable changes in cataloguing rules, classification schemes, machine formats, and subject heading lists. Perhaps more should do so.

What are the benefits of new cataloguing rules? Insofar as they deal with new conditions and embody new theories, they provide a basis for co-operation in the modern era - an era in which shared cataloguing is the lifeblood of most libraries. They provide, at least in the case of the cataloguing rules published in the last decade, a more user-oriented bibliographic system. Further, new cataloguing rules which meet current conditions provide cataloguers with a congenial working framework. That framework creates a more productive and efficient working environment, especially when it is contrasted with the existential angst produced by trying to fit new materials into an outmoded set of rules. Lastly, they provide a reason, a pretext, or a tool by which the whole bibliographic environment can be changed. That change, which is inevitable anyway, is from old technologies of

bibliographic control to new computer-based systems. It is certainly not always necessary to close the catalogue in order to adopt new rules (in this context, see Taylor Dowell's excellent article[13]). However, if the catalogue is closed and a new catalogue in a new form is created, the benefits can be maximized and the costs can be reduced to a minimum. In fact, a new catalogue based on new rules can be more cost-efficient than the old cataloguing system.

There is a present need for a continuing alliance between library administrators and bibliographic experts. There is a present need for continuing communication between the two groups. This alliance and communication would be dedicated to the management of change in bibliographic systems and to the reconciliation of bibliographic imperatives and financial exigencies in order to achieve the co-operation and user-friendliness which we all desire.

REFERENCES

1. Lubetzky, Seymour. Cataloging rules and principles. Washington, D.C.: Library of Congress Processing Dept., 1953.

2. International Conference on Cataloguing Principles, 1961, Paris. Statement of principles. Annotated. ed. London: IFLA Committee on Cataloguing, 1971.

3. Gorman, Michael. "International Standard Bibliographic Description and the New ISBDs." Journal of Librarianship, 10, April 1978.

4. Domanovsky, Akos. Functions and objects of author and title cataloguing. Budapest: Akad. Kiado, 1974.

5. Verona, Eva. Corporate headings. London: IFLA Committee on Cataloguing, 1975.

6. Kaltwasser, F.G. Universal bibliographical control. Unesco Bulletin for Libraries, v. 25, p. 252-259, Summer 1971.

7. Anglo-American cataloging rules. North American text. Chicago: A.L.A., 1967. (AACR1)

8. Swanson, G.L. ISBD, standard or secret? Library journal, v. 98, p. 124-130, Jan. 15, 1973.

9. AACR 1. Chapter 6, revised. Chicago: ALA, 1974.

10. Anglo-American cataloguing rules. 2nd ed. / edited by Michael Gorman and Paul Winkler. Chicago : ALA ; London : Library Association ; Ottawa : Canadian Library Association, 1978.

11. Martell, Charles. War of AACR2. Journal of academic librarianship, v. 7, pp. 4-8, July 1981.

12. Gorman, Michael. And now, from those wonderful folks who gave you superimposition ... American libraries, v. 9, p. 620-621, Nov. 1978.

13. Dowell, Arlene Taylor. Impact of AACR2. Library research, v. 3, p. 107-122, Summer 1981.

163-CAT-4-E

Project ESSCAPE -
Expert Systems for Simple Choice of Access Points for Entries:
*Applications of Artificial Intelligence in Cataloging**

R Hjerppe, B Olander†, K Marklund‡

LIBLAB

June 1985

Linköping University

Department of Computer and Information Science and University Library

Meeting 123

*Paper for presentation at IFLA 51st General Conference in Chicago, 18-24 August 1985. This paper is a condensation of LiU-LIBLAB-R-1985:1

†Also at University of Toronto, Faculty of Library and Information Science, Canada

‡ Lund University Library, Sweden

1. LIBLAB

LIBLAB is a joint project of the University Library and the Department of Computer and Information Science at Linköping University in Linköping, Sweden. The project has been initiated with the twofold aims of 1) creating a basis for continuing library research, which until now has not been available in Sweden, and of 2) pursuing research in some of the areas central to libraries.

Funding of LIBLAB for the initial three years, starting January 1, 1983, was provided by the Swedish Delegation for Scientific and Technical Information, DFI, (a body corresponding to NCLIS in the US), on the basis of an application and a research program presented in a report (1).

2. RESEARCH AT LIBLAB

The research program for LIBLAB outlines two major and one minor themes, proposed as foci for the research of LIBLAB. The central theme is Document Description and Representation, and the other major theme is Users and Library (Systems). The minor theme is Networking.

For the first years the central theme has been further delimited to cataloging and cataloging rules. A first attack has been made in this area with expert systems as the "weapon". Some of the results from that, and from ensuing research, will be described in this report.

An initial survey has also been made of the various forms of bibliographic respresentations handled in libraries and the transformations of these representations (2).

During 1984 work has also been initiated on a project called "HYPERCATalog", concerned with the structure and design of an enhanced catalog, that would contain more and other information than present catalogs, especially in terms of links and relations between fields, records and files, and facilities for utilizing this. The HYPERCATalog, as envisioned, will support browsing and navigation as the primary modes of using the catalog in addition to traditional searching, in which the user has to specify what he wants. Furthermore each user will be able to specify, and save, his views of the library. One of the goals of the project is to make the catalog dynamic, enabling it to be enriched by the use that is made of it (3,4,5).

During 1985 an attempt will also be made to develop a formal model for the description of documents, based on a format for abstract editors.

3. ARTIFICIAL INTELLIGENCE AND LIBRARIES

Artificial Intelligence, AI, in the last few years has received increasing attention in the library and information service community. A good starting point for reading more about AI and libraries and information services in general is the review article by Linda C. Smith in ARIST (6). A number of later papers discussing mainly the potential uses of AI in libraries and information retrieval services are e.g. (7-10).

In the future libraries are likely to use AI applications for various purposes. At present there is considerable effort spent in developing "intelligent" information retrieval

systems, and operational "semi-intelligent" systems exist today. To be called "intelligent", a retrieval system should be able not only to fully comprehend the query and to retrieve explicit information from documents, but also to extract the implicit information in them and to provide answers using that information. User interfaces with natural language processors, and inference functions for answering questions are AI applications that generally would seem to be of special interest to libraries.

4. EXPERT SYSTEMS AND CATALOGING

Expert systems are perhaps the result of research in AI that has received the most attention recently. Very briefly expert systems can be characterized as systems in which the knowledge of an expert in a specific problem domain has been captured and can be manipulated. Hence, expert systems provide access to such knowledge. The manipulation of it includes deductions, i.e. proposed solutions/answers, as well as explanations. Descriptions of expert systems can be found in e.g. (11,12). Expert systems are usually looked upon as one type of knowledge-based systems.

A number of turnkey skeletal expert systems are now available on the market. The problem solving stategy is "built in", and the user sets up his own problem domain on the frame offered by such a system (13).

Most of the attention on expert systems in the library and information service area has focussed on assisting users in online information retrieval services. To our knowledge the first attempt at using expert systems for activities related to traditional library tasks was at the University of Exeter, where work on an expert system for the rules in Chapter 21, Choice of Access Points, of The Anglo-American Cataloging Rules, 2. ed. (AACR2), was carried out between October 1982 and October 1983 (14,15). Nothing had been published about the Exeter project at the time we initiated ours so we were unaware of their activities. The two studies started independently of each other but contact is now established, and we have found that our experiences and insights in many essential respects are similar, although different techniques were used.

There have been earlier attempts at formalizing the application of cataloging rules, however. "Using AACR2: A diagrammatic approach", (16), first published in 1980, (which was used at the University of Exeter) is one such example. Another is the dissertation "The amenability of a cataloging process to simulation by automatic techniques" by Ann Fox (17), in 1972. Both are concerned with the choice of access points. In a recent paper AI and authority control is discussed (18).

5. THE PROJECT, ESSCAPE

The project, Expert Systems for Simple Choice of Access Points for Entries - ESSCAPE*, described in this paper, does not have as a goal to produce an operational system to be used in actual cataloging. It is rather an open-ended, continuing process in which the applicability of various tools to the handling of sets of rules is tested, and in which the set of rules and their application and use are elucidated by this testing.

---

*"*escape: the departure from one code or language to another code or language, i.e. the withdrawal from existing pattern.*" (Computer Dictionary, by Sippl & Sippl, 3rd. Ed. 7th pr. 1984)

The set of rules used in ESSCAPE is AACR2, and so far mainly Chapter 21, Choice of Access Points. It was chosen because it is suitably complex (it is often perceived as the most difficult part of the catalog code to apply), it is decomposable and interpretations in a formalized form are available.

For the moment the approaches reported here are not being pursued further. The project has now taken a different direction. A first version of an experimental system has been developed for authority control of personal names, applying rule-based (AACR2, Ch. 22, Headings for Persons) transformations on the input to a name database. The problems of names and uniform headings, disregarded in the work presented here, are the next area to be studied within the ESSCAPE project.

5.1 Motives for the project

Document description and representation is the major research theme that LIBLAB is now concentrating on. For the moment it has been further delimited to the descriptions made and the representations produced in cataloging and the uses of these in catalogs. Cataloging is based on a set of rules and hence these have become the object of our investigations.

The main reason for this choice is that for quite some time libraries have been in a process of transition, from traditional manual methods of information handling to automated computer-based ones. So far that transition has been characterized by attempts to do the same things as before, only by different means.

It was apparent as early as 1968, two years after the MARC-project was initiated, that *"A machine record is not simply a different physical means of recording a traditional bibliographic entry, for use in a traditional way"* (19), but the library world by and large has not heeded the accompanying plea for research aimed at the determination of the structure and content of an optimal bibliographic record (in machine-readable form).

The only ones to experiment and utilize the inherent possibilities of machine-readable bibliographic representations have been the abstracting and indexing services, mostly more concerned with costs and benefits than libraries. The advent of Online Public Access Catalogs is nevertheless making librarians realize that there are other presentations and uses of bibliographic representations than those of the card catalog.

Furthermore, however sophisticated and "user-friendly" systems we might design, build and implement, they will be of no avail if the data, the representations, they manipulate are found wanting in structure and content. Hence there is a need to learn from the erudition of the past, the experience and mistakes of present practices and experiments of information services, using fully the potential capabilities of information technology and the theories of computer and information science.

One of the issues facing us in this situation was methodology. How does one start analyzing a set of prescriptive rules? What would characterize a "bad" vs. a "good" set of rules? What tools are available?

The method chosen was to use knowledge-based systems and try to embed the given rules in such systems. This is admittedly an unorthodox application of these systems but it also does provide for some interesting insights into the uses and functioning of them.

Within LIBLAB there were a number of somewhat different goals with this approach:

- To get a thorough acquaintance with expert systems and artificial intelligence, i.a. for assessment of its impact on and relevance to libraries.
- To develop demonstrable "prototypes" for the library community, one cataloging consultant for librarians and another for catalog users.
- To try to produce a simplified, reduced set of rules compatible with the complete rules for all those who only occasionally do cataloging, and eventually to be able to start devising completely new, better rules, to make a small start on AACR3.
- To look at an application of expert systems with rules as a domain. This is of special interest since most applications so far have been on descriptive rules rather than prescriptive.
- Methodologically to use a combination of experiment and reflection, an ability to test and try both tools and their objects.
- Last but not least important, testing the hypothesis that expert systems would provide a powerful instrument for revealing the structure and nature of some aspects of cataloging and cataloging rules.

5.2 Approach and tools

Our basic approach in embarking on this venture can best be described:

From very simplistic to somewhat more complex.

The "tools" used by us were first EMYCIN, a system for building expert systems on DEC-20 mainframe computers (20), and later Expert-Ease as implemented on a LISP-machine, Xerox 1108, a powerful workstation (21).

5.2.1 ESSCAPE/EMYCIN

The actual design and implementation of ESSCAPE/EMYCIN was made by Birgitta Olander, a librarian working at LIBLAB, with no previous formal training in computer science (an indication of the "relative ease" of building an application with EMYCIN).

Within EMYCIN knowledge is encoded as so-called production rules:

If *condition* **then** *action*

EMYCIN is a goal-driven, backward-chaining expert system. The inference strategy used by such systems is to apply the rules of its knowledge-base in order to achieve goals that have been set up as hypothetical solutions to the problems of its domain. In order to establish the status of the goal parameter, the system works "backwards" through the inference rules set up by the programmer, that are known to bear on this goal. For ESSCAPE/EMYCIN the overall, top-level goal is to find out whether the bibliographic record is completed or not. A forward-chaining system, on the other hand, would start with low level data and systematically evaluate all the alternatives that they give rise to.

In ESSCAPE/EMYCIN a complete record is defined as containing a complete bibliographic description as well as appropriate access points. (In App. 1 there is a general presentation of ESSCAPE/EMYCIN). If the system cannot decide the status

of the record by just using the rules, it has to ask the user for more information. In actual practice the user is prompted for the piece of information needed by the system for its inference process, e.g. the elements of the bibliographic description. Superficially this is similar to a menu-driven input system for cataloging.

In the very first version of ESSCAPE/EMYCIN (of which there are examples of a number of consultations in App. 1a) it was assumed that a description had already been made. Demonstrations of this version proved, however, not to be very satisfying although the system performed as planned. The eliminated description process was apparently necessary to provide the proper context and to relate to the ordinary cataloging situation.

In the present version the cataloging process is modelled so that a simple bibliographic description is first prompted for. When the system recognizes that the description is complete according to the criteria defined, i.e. the first of the two subgoals necessary to attain the top-level goal is reached, the process goes on to establish the access points in order to reach the second subgoal. (In App. 1b there are examples of a number of consultations using this version).

5.2.2 ESSCAPE/Expert-Ease

Expert-Ease is the name of a simple system originally developed by Prof. D. Michie's Intelligent Terminals Inc. for IBM PCs and compatible systems. Expert-Ease does not meet all the criteria normally set for expert systems - it does not have any explanation capabilities, it lacks a proper knowledge base, etc., and ore should not be compared to e.g. EMYCIN - but might still be useful for certain applications. One of its merits is that an application is built by giving a number of examples, from which a generalization is made by the program. Thus the need for formalization of the input is not very high.

With Expert-Ease the application is first described in terms of attributes of the objects, the possible values of those attributes, and expected results. Then a number of examples, with varying values for the attributes, are given together with the corresponding outcome for each example. When the number of examples is considered satisfactory, Expert-Ease can be used to generate and display a decision tree for the application domain.

In building the system matrices of an electronic spreadsheet type are used to define the application. In using the system for consultations it will ask questions and the user responds by selecting from menus provided by the system.

In App. 2 there is a general presentation of ESSCAPE/Expert-Ease, and in App. 2a and 2b there are two examples of consultations from the different versions of ESSCAPE/Expert-Ease that have been produced. The first example, App. 2a, corresponds to the present version of ESSCAPE/EMYCIN, and the second, App. 2b, is an application of a flowchart of the major decisions in AACR2 made by Cockshutt et al. (22), and displayed in App. 3.

5.3 Results and revelations

The goals described earlier have not all been reached as yet. The results so far will be described under two headings:

- systems produced, i.e. ESSCAPE-systems,
- observations on AACR2.

The revelations, i.e. things revealed or insights, that will be discussed are:

- expert systems for cataloging,
- cataloging as interpretation,

The presentation will be a mingling of these, in order to relate results and insights to each other. Furthermore the results and revelations are only briefly indicated in this paper. Extended discussions of specific items will be forthcoming in the LIBLAB report series.

5.3.1 The traditional catalog

Our results and revelations are presented within a frame based on the functions and form of the traditional catalog, as found in e.g. card catalogs or online public access catalogs, OPACs (that are actually modeled on the card catalogs).

Traditionally the functions of a catalog are stated as:

- to show whether a specified item is in the collection,
- to bring together related items,
- to indicate the physical location of the items in the collection.

Those functions are achieved basically by descriptive cataloging. A fourth function, enabling the user to find items by subject content, is achieved by subject cataloging. Usually those two types of cataloging are regarded as two separate processes.

In the rest of this paper we are concerned only with the author-title catalog and descripitive cataloging. The words catalog and cataloging should be read author-title catalog and descriptive cataloging, respectively, unless otherwise specified.

Traditional cataloging can best be described in terms of bibliographic work for the production and maintenance of a card catalog. From this viewpoint cataloging is the process whereby

- an item is bibliographically described according to predefined rules,
- access points for the item are chosen according to predefined rules,
- the form of each access point is determined, and
- a representation of the item (description), or a pointer to one, is filed under each access point.

Underlying this is the view of the catalog as a simple linear file, where the choice of access points predetermines the ways in which one will be able to retrieve the representation, and where the ordering of the file is concomitant with the need for determining the form of each access point.

5.3.2 Results: ESSCAPE-systems

Two ESSCAPE-systems, each in several versions, have been produced as mentioned earlier and as demonstrated in appendices 1-3. The systems produced provide different

types of advice, ranging from pointers to relevant rules in AACR2 to simple complete records. Both systems handle only a small part of the rules in AACR2.

Both systems are designed for the cataloger. No system has as yet been produced that attempts to assist the patrons in using the library catalogs.

Some of the differences between the ESSCAPE/EMYCIN and ESSCAPE/Expert-Ease systems are the following:

- A session with ESSCAPE/EMYCIN is fairly similar to cataloging with an operational system, in which the system cues the user for input and a catalog record is produced as a result of the consultation. A session with ESSCAPE/Expert-Ease will only result in an indication of the rules that should be consulted.
- The communication with the user in ESSCAPE/Expert-Ease is very brief, almost cryptic, whereas it is verbose in ESSCAPE/EMYCIN.
- The rules have to be translated to "if condition then action" form in ESSCAPE/EMYCIN, whereas they have to be transformed to examples in ESSCAPE/Expert-Ease.
- ESSCAPE/Expert-Ease was easily given a wider scope in terms of number of rules covered, although it became more tedious as the scope grew. The initial effort was greater in ESSCAPE/EMYCIN.
- Decision flows that are not strictly hierarchical, i.e. in which a decision is later neglected, can only be implemented in ESSCAPE/Expert-Ease by repeating all the examples with a variation in these parts, which also means that the decision tree gets repeating branches. This situation does not create any problem in ESSCAPE/EMYCIN.

5.3.3 Revelations, expert systems for cataloging

The revelations in this section are our opinions on the technical feasibility of building expert systems for cataloging based on our results and experiences so far. There are also other, more important factors to be considered that will be discussed in other sections.

All of AACR2 is probably too large to have in one system at present, but then nobody needs to use all the rules at the same time. One solution might be to use separate expert system modules for separate types of documents to be called upon as needed. Such modules could be built from submodules since the goal of the structure of AACR2 is to distinguish the general framework from its application to specific types of material.

An integrated system that is restricted to one type of material but using all relevant rules, i.e. an expert system having as its knowledge base the preface, the introductions, Chapters 1-2, 21-26, and the appendices, would probably stretch the limits of the tools available at the moment. An added complication is the fact that the rules are not self-contained but references to other information sources abound, especially in chapters 22 - 25, and Appendix 1. Hence, any expert system for cataloging aiming at a level of performance corresponding to that of an expert cataloger would also need access to an authority file and/or a catalog.

A restricted system, limited to description and choice of headings, i.e. an expert system without Chapters 22-26 and appendices, is technically feasible, but would be cumbersome because of problems with contradictions and lack of clarity in AACR2. This will be further discussed later in this paper.

Traditional cataloging is seen as two separate sub-processes: bibliographic description, and identification of access points. The distinction of those two parts of the cataloging process permeates both AACR2 and the MARC-format. An expert system must have the ability to handle and integrate both parts of the process, although they require very different skills and use different problem solving strategies.

5.3.3.1 Main entry heading

In a computer-based catalog there is no intrinsic need to separate access points from descriptive elements, i.e. to designate some of the parts of the bibliographic record as access points and others as descriptive elements. Every separate element of the record that is identified as a field or subfield of its own may in principle be an access point. In actual practice, not all elements in the record are made available as access points because of economic limitations as well as limited usefulness.

However, as soon as we wish to display several entries, the need arises to order or sort them. Out of this need comes the necessity to select one descriptive element as the main sorting key. Such a key might be called a main entry heading, even though it does not necessarily correspond to the traditional concept of such a heading.

The need for choice of access points in the traditional sense thus is not present in the online catalog. On the other hand there is an increased need for enhancing the record with more or new elements that improve access to the item.

5.3.3.2 Authority control

It is necessary to identify the appropriate form of main entry headings, and to use that form consistently in the catalog. The process of establishing a preferred version and form of names, personal and institutional, is called authority control, because the version and form chosen are considered authoritative thereafter.

There may be problems similar to those of names with the identity and physical manifestations of works. The traditional solution is to establish a preferred version and form of a title, called uniform title, to link different manifestations of the same work.

The traditional catalog by itself might be a tool for the maintenance of such authority control, or a separate system can be used. Either needs, however, to be supplemented by other information sources as soon as a new name or title is encountered.

Similar problems are encountered in describing the content or subject of the items in the catalog. A classification system combines three functions, subject structuring, authority control, and collocation.

There are other solutions to such general authority control problems in a computer-based catalog than the collocation device used in traditional printed catalogs. The most obvious solution is simply to make all variations and versions and their forms equally valid and provide links between them. In a generalization and

extension of this principle any part of the bibliographic record, including elements that are not traditionally bibliographic, might be linked.

With a slightly pedantic formulation we might claim that the traditional choices of access points and the determination of their form is a kind of description, one that imposes on the item descriptive elements or forms that are not necessarily found in the item itself.

5.3.4 Revelations, cataloging as interpretation

Interpretation is based on recognition and identification. Much of the present cataloging process consists of unconscious interpretation based essentially on experiential learning from examples, in an apprenticeship manner. Hence cataloging as a craft rather than as an art. "The interpretative art of cataloging" (23, p. xi) is a recognised concept, but the different types of interpretations performed and the full extent of each of these seems to have received little attention.

More important, the consequences of such interpretations for cataloging theory, rules and practice appear not to have been considered. The discussion to follow will again be confined to regarding the two prime traditional aspects of cataloging, description and choice of access points.

5.3.4.1 Description and interpretation

Description, and the purpose of description in cataloging, has received much less attention than the choice of headings in debates over cataloging. Until recently the cataloging rules were almost completely neglecting the descriptive part of cataloging, centering on the access point problem, and especially the main entry heading. The objective of bibliographic description is very little discussed, cf for example a definition like: *"... descriptive cataloguing: the identification and description of a work in such a fashion that it can be distinguished from all other works, and from other editions of the same work. As Paul Dunkin put it, this is the part of the catalogue entry that tells what the work looks like - rather than who is reponsible for it."* (24, p. 9).

Description is, however, neither simple nor innocent. All description is based on a complex composite of perception, recognition, and selection of characteristics. The results of this process are presented symbolically, in a highly formalised bibliographic record. The structuring as well as the actual words carry information for those users who share the frames of reference with the cataloger.

The efforts of implementing the cataloging process in a computer system elucidate the complexity involved in the process of reading and correctly interpreting the symbolic information of a title page. In this way we can get an indication of the extent of competences required and interpretations performed in the cataloging process that we are usually not explicitly aware of.

> Assuming that we have a system that in a limited sense is able to "see" and "read" a document, we find that the very first obstacle is in defining the chief source of information, the title page. Attempting to use the definitions in AACR2 we run into circular reasoning.

- "*Title page. A page at the beginning of an item bearing the title proper and usually, though not necessarily, the statement of responsibility and the data relating to publication. ...* " (App. D Glossary, AACR2)
- "*Title. A word, phrase, character, or group of characters, normally appearing in an item, naming the item or the work contained in it. ...* " (App. D Glossary, AACR2)
- Prescribed source of information for title and statement of responsibility area is title page. (Rephrasing of Table in 2.0B2, AACR2)

Attempting to define a title page by a set of criteria we find that there is no simple list of criteria that is satisfactory since the latitude for variation is so great. Obviously the layout of the title page plays an important role, but only in combination with other factors. Many books have e.g. added title pages with similar types of layout, and sometimes the information usually found on the title page is spread out on several pages. In addition to being able to use the clues provided by the layout the system must be able to identify names and to differentiate between personal, institutional, and geographical names.

The system would have to have the ability not only to correctly identify a title page, but also to categorize all the text that is not names, decide upon the relations between the names and other elements of the text, etc. It would need competence for interpretation of the function of some names as indicating authorship, others as publishers, etc..

Recognition of names is important, and one of the really difficult tasks would be to formulate criteria for distinguishing between title and author. Names probably are less difficult to recognize than titles, but many titles consist solely of names.

Our experiences from the ESSCAPE project very clearly indicate that interpretation permeates the descriptive part of cataloging to a much larger extent than has been recognized before.

Therefore it is not surprising to find the following statement in the preface of a textbook on cataloging (24 p.ix):

"*Furthermore, it has been the author's experience in many years of teaching cataloguing that one of the most difficult concepts for beginning cataloguers is the translation (sic!) of the title page (or other chief source) into a catalogue entry*"

The solution usually adopted in teaching cataloging is to provide liberally copies or transcriptions of a wide variety of title pages and other sources of information. Learning cataloging thus is an acquisition of tacit knowledge by studying examples, at the same time as it is a course in the variability of title pages and their interpretation.

5.3.4.2 Choice of access points and interpretation

Choice of access points cannot be separated from description when applying AACR2. Decisions (interpretations) made in the description process delimit the choices of access points.

When looking at choice of access points as interpretation it becomes apparent that there are four different issues that need to be emphasized:

- authorship,
- relationship (if any) to other items,
- authority control, and
- choice of access points as a provision of access to the item.

The establishment of responsibility for the creation of the item, i.e. authorship, actually is an interpretation of the title page, its layout, the fonts used, the use of prepositions or phrases, etc. This interpretation is used for ascribing degrees of responsibility to various persons/institutions whose names appear on the title page or its equivalent.

The establishment of relations (if any) to other items is likewise an interpretation, based on comparisons of two or more items. The primary result from such a comparison is a decision whether the item should be looked upon as changed or not, with regard to appearance and/or content.

Authority control, "uniforming" the access points chosen, also implies a judgement, a choice of what variation of name to use and its appropriate form.

All those interpretations are being made because of the rules that govern choice of access points. The Paris Principles (25) designate the principal author's name, or the author named first on the title page, as the main entry heading, and specify that an entry be made under a uniform heading, if there are variant forms of the name. Out of such decisions arise the needs for ascribing degrees of responsibility, and for uniforming the heading, as well as judging relationship to other items.

5.3.4.3 Types of interpretation in cataloging

Looking at cataloging in a broader perspective we find that there are a number of rather different types of interpretative acts performed:

- Item / Description related:
  - recognition of, and distinction between, the types of material cataloged,
  - recognition of the basic types of entities used in bibliographic description, viz. title page, names, etc.,
  - interpretation of the function of the entities recognized, and
  - tentative transcription, and recording of them.
- Item / Access point related:
  - interpretation of function of layout, fonts, etc., for
  - interpretation of degrees and kinds of responsibility for the creation of the item, for
  - ascribing principal responsibility to one entity, for use as main entry heading, and secondary or other type of responsibility to other entities for added entry headings, or establishing no single principal authorship, implying use of the title for main entry heading,
  - recognition of name or title as simple or "exceptional",
  - recognition and interpretation of parts of name, or title,
  - if necessary, transformation of name or title to uniform form,

- recognition of item as possibly being related to other item(s).

Cataloging rule related:

- interpretation of the cataloging code as a whole, and the relations between general and specific rules,
- interpretation of specific rules and examples, in the context of the rule set, and related rules,
- selection of relevant rule(s),
- preliminary interpretation of the rules,
- re-examination of item and entities identified,
- interpretation, and rejection or application of rule(s).

This list is not intended to be a definitive model in the sense of implying that the various interpretative acts performed by a cataloger are actually those listed, or that they follow a simple sequence. On the contrary it seems clear that in cataloging the intepretative acts are highly interrelated, quite probably even in a circular, i.e. iterative manner.

5.3.5 Results, observations on AACR2

Most of the preceding discussions are generalizations from experiences with and understanding of AACR2 gained in the research on the ESSCAPE-systems and on related subjects. In this section some additional observations on AACR2 will be discussed.

5.3.5.1 AACR2, a critique

The brief critique below of AACR2 is a selection of some observations of problems that have not been discussed previously in this paper.

There is no presentation of what the catalog is on a conceptual level, only of what its function should be (according to Cutter, and as later raised to canonicity by the Paris Principles). The catalog is seen as a tool for achieving the two (perceived as conflicting) goals of providing access by author/title on the one hand, and linking related works on the other.

There is no explicit recognition of the fact that the present rules are built on an implicit conception of the catalog as a simple linear file, and the catalog record as something confined by the limitations of a 3" x 5" card.

AACR2 is self-contained, but not self-supporting. Self-contained in the sense that the only reference made in part I to other texts concerned with bibliographic description is to ISBD(G). Not self-supporting because it cannot be used as it is, without a course in cataloging, and because it does not recognize the problems of interpretation indicated above.

Like most professional handbooks it is written for those who already know.

Finally, some general observations:

- The basic entities are not properly defined.
- Consequently, concepts are overlapping and rules badly structured.
- Alternatives are not exclusive, or precedence rules not formulated.
- Distinctions and exceptions seem to have been made on an ad-hoc basis.
- The overall structure is fairly well conceived but mostly poorly executed.
- The relations between general rules and specific are not sufficiently understood.

6. CONCLUDING REMARKS

Though e.g. Lubetzky, Ranganathan and Cutter have provided insights, distinctions and useful precepts, much of the literature on cataloging has consisted of itemization and theorizing, attempts to cover every conceivable situation with a special rule. The same affliction still plagues much of cataloging, although in different form. The attempt is now instead to try to force all kinds of collectibles into one descriptive straitjacket, modeled on books, and even worse, to enforce the concept of assigning responsibility for the creation of items (authorship) to persons or institutions irrespective of the type of item, with one main "author" as the normal case.

6.1 Towards a theory of catalogs

There is still no general, formal theory for catalogs. A general theory for catalogs would provide at least a basis for the discussion of issues such as those in the following, very incomplete list:

- The nature of the collection as a determinant of the characteristics perceived and selected for description in the items collected, or rather the interdependence of the nature of the collection and the perception of characteristics of items.
- The relationship between description of items and the use of descriptions, for identification and differentiation of items, for access to items, for judging items, for structuring collections.
- The functions of and relationships between various characteristics of items, e.g. those "on" it, those "in" it, those inferred or deduced from it, those ascribed to it, etc.
- The catalog as a representative of the collection but with characteristics of its own, as well as of the collection, not only as a set of representatives of items.
- The relationships between the form/material of the carrier of the representation of the item, and the representation, and the form of the catalog, the technology used, and its structure, etc.
- Measures of catalogs, their structure, complexity, etc.

6.2 Expert systems for cataloging

Answers to questions that center on the true expertise of an expert cataloger, his/her tacit knowledge, are essential for the design of operational expert systems for cataloging assistance. Such questions might be:

How does one recognize anomalies, when to apply which rules? What clues are used in detecting, or rather forming a hypothesis on, exceptions from the "normal" cases? What procedures are followed, what information is sought, and what is used in testing the hypotheses? How much are the rules used in comparison with, and in cooperation with, reference to personal examples, earlier cases and their solutions?

The role of examples is stated in AACR2 as follows: "... *they are intended to illuminate the provisions of the rule to which they are attached, rather than extend those provisions.*" (Rule 0.14). Their role in an expert system for cataloging, and their relation to local practice which has to be essential for expert system applications, is something that needs to be investigated.

Most of the questions above refer to the acts of interpretation described above. Such acts require personal experience and skill from the cataloger, i.e. what might be called expertise, as contrasted to the formally expressed knowledge of the cataloging rules.

We have to bear in mind that the aptness, in contrast to "correctness", of a catalog record can be judged only in its local environment. AACR2 is a frame of reference common to all libraries that use it, but it still leaves ample room for local praxis and individual discretion, and rightly so. Thus, an expert system for cataloging based on AACR2 in order to be truly useful has to be a tool designed for and applied to its own local environment. The relation of such systems and their products to national bibliographies and their contents is an issue for future deliberations.

Appendices:

1. A general presentation of ESSCAPE/EMYCIN by B. Olander
1a. Examples of sessions with ESSCAPE/EMYCIN version 1
1b. Examples of sessions with ESSCAPE/EMYCIN present version

2. A general presentation of ESSCAPE/Expert-Ease by R. Hjerppe
2a. Examples of sessions with ESSCAPE/Expert-Ease version 1
2b. Examples of sessions with ESSCAPE/Expert-Ease version 2
3. Overall decision flow in chapter 21 of AACR2, as described by M. E. Cockshutt et al.(22)

REFERENCES:

1. *Hjerppe, R.; Bivins Noerr, K. and Noerr, P.:*
   LIBLAB Research Program
   (Linköping Univ. Rep. LiU-LIBLAB-R-1982:3 Sept. 1982, 44p.)

2. *Gustafsson, B.:*
   En studie av bibliografiska representationer i LIUBs förvärvsregister.
   (LiU-LIBLAB-WP:27 Febr. 1985, 11p.)

3. *Hjerppe, R.; Järvelin, K.:*
   Projekt HYPERCATalog.
   (LiU-LIBLAB-WP:9 Jan. 1984, 12+15p.)

4. *Hjerppe, R.:*
   HYPERCATalog projektet. Några förtydliganden.
   (LiU-LIBLAB-WP:16 May 1984, 4+12p.)

5. *Hjerppe, R.:*
   HYPERCATalog projektet. Ytterligare förtydliganden och reviderad budget.
   (LiU-LIBLAB-WP:18 Sept. 1984, 5+1p.)

6. *Smith, L.C.:*
   Artificial Intelligence Applications in Information Systems.
   (Ann. Rev. Inf. Sc. Tech. (1980) vol.15 pp. 67-105.)

7. *Davies, R.:*
   Documents, information or knowledge? Choices for librarians.
   (J. Librarianship (1983) vol.15 no.1 pp.47-65)

8. *DeJong, G.:*
   Artificial Intelligence Implications for Information retrieval.
   (Proc. 6th Ann. Int. ACM SIGIR Conf. on R in IR. ACM SIGIR (1983) vol.17 no.4 pp. 10-17)

9. *Clarke, A., Cronin, B.:*
   Expert systems and library/information work.
   (J. Librarianship (1983) vol.15 no.4 pp.277-292)

10. *Hjerppe, R.:*
    What artificial intelligence can, could, and can't, do for libraries and information services.
    (LiU-LIBLAB-R-6, Sept. 1983, 20p.) (Also published in Proc. 7th IOLIM, Dec 6-8 1983, Learned Information Ltd. London.)

11. *Nau, D.S.:*
    Expert Computer Systems.
    (Computer (1983) vol.16 no.2 pp.63-85)

12. *Gevarter, W.B.:*
    Expert systems: limited but powerful.
    (IEEE Spectrum (1983) Aug. pp.39-45)

13. *Hayes-Roth, F., Waterman, D.A., Lenat, D.B., Eds.:*
    Building Expert Systems
    (Addison-Wesley Inc., London, 1983, p. 286)

BEST COPY AVAILABLE

14. *James, B.:*
Expert Systems for Library-Cataloguing Functions.
(M. Phil. Thesis Fac. Sc. Univ. Exeter, Oct. 1983, 119 p.)

15. *Davies, R., James, B.:*
Towards an expert system for cataloguing: some experiments based on AACR2.
(Program vol.18(1984)n.4 pp.283-297)

16. *Shaw, M.; Dent, B.; Evans, D. & Smith, D.:*
Using AACR2: A Diagrammatic approach.
(Oryx Press Phoenix 1981)

17. *Fox, A. M. S.:*
The amenability of a cataloging process to simulation by automatic techniques.
(Ph.D. Diss. Univ. Ill. Urb.-Ch. 1972 476p.)

18. *Burger, R.H.:*
Artificial Intelligence and Authority Control.
(Libr. Res. Techn. Serv. Oct./Dec. (1984) pp.337-345)

19. *Vickery, B.C.:*
Bibliographic Description, Arrangement and Retrieval.
(J. Documentation vol.24(1968)no.1)

20. *Melle, W.; Scott, A.C.; Bennett, J.S. &Peairs, M.A.S.:*
The EMYCIN Manual.
(Comp.Sc.Dept. Stanford Univ. Report No. STAN-CS-81-885 Nov. 1981 138p.)

21. *Moen, S.:*
En implementering av Expert-Ease under Interlisp-D.
(Dept.Comp.Inf.Sc. Linköping Univ. ASLAB MEMO 84-04, Oct. 1984 21p.)

22. *Cockshutt, M.E, Cook, C.D., Schabas, A.:*
Decision logic for Anglo-American Cataloguing Rules, Chapter 21: "Choice of Access Points".
(Libr. Res. Techn. Serv. vol.27(1983) pp.371-390)

23. *Wynar, B.S.:*
Introduction to Cataloging and Classification.
(Libraries Unlimited, Inc., Littleton, Col. 1980)

24. *Maxwell, M.:*
Handbook for AACR2.
(American Library Association, Chicago 1980)

25. *International Conference on Cataloguing Principles Paris 1961*
Statement of Principles.
(IFLA Committee on Cataloguing, London 1961)

IFLA GENERAL CONFERENCE, CHICAGO 1985

Division: Bibliographic Control

Section: Classification and Subject Cataloguing

47-CLAS-1-E

1985

Classification in the USA

by

Robert P. HOLLEY
Assistant Director for Technical Services
University of Utah Libraries
Salt Lake City, Utah, U.S.A.

For internal use only

To be brought to meeting on:

. . . . . 87 . . . . . .

Remarks:

to be presented orally ☒

for distribution only ☐

BEST COPY AVAILABLE

R. Holley (USA). Classification in the USA

American libraries value classification as a means for subject browsing in open stacks. Dewey Decimal Classification, devised as a theoretical attempt to organize all knowledge, is used by 85% of American libraries including most school and public ones. Library of Congress Classification lacks intellectual consistancy since it was based upon literary warrant as a pragmatic classification to organize materials at the Library of Congress. Many academic libraries use L.C.C. because it reflects their collecting scope and because LC provides its call numbers in its shared bibliographic records. L.C.C. is more hospitable to change than D.D.C. which has encountered resistence to its policy of reassigning numbers through Phoenix schedules. Though perhaps in time more important for international subject access, classification currently receives less attention than subject headings because American librarians place great hope in the computer to resolve subject heading problems while remaining conservative about classification.

CLASSIFICATION IN THE USA

Today's speech posed many problems to me during the preparation period.[1] First, how can I even begin to summarize the past, the present, and the future of classification in a country with 29,044 libraries?[2] Second, how should I develop this topic at an international forum? I must provide enough background material for our visitors who may not be familiar with librarianship in the United States while yet not boring my American colleagues. Third, could I free myself enough from my own experiences as an American librarian to achieve an international perspective?

As my solution, I will try to explain classification in the American context and to illustrate how the theory and practice of classification grew out of certain principles of American librarianship. I also apologize in advance if my perspective at times appears naive because I lack information on classification in other parts of the world. This lack in itself might be a comment on the American character because we have not imported foreign classification schemes to any great extent and are in general deficient in our understanding of theoretical work done elsewhere. Finally, I have been guilty of rash generalizations in the past and will surely be guilty of a few more today.

As I see matters, significant factors in the American way of classification have been:

1. Open stacks. American libraries of all types have a firm tradition of open stacks. While this factor could be attributed to a desire to allow patrons to retrieve their own materials after finding them in various bibliographic sources, the evidence does not support this conclusion. Open stacks do not require, in principle, classification since patrons could retrieve books by means of an accession number. In fact, shelving library materials by type and by size with an accession number is a practical solution because it makes more efficient use of buildings; sequential accessing does not require unpredictable growth space, and shelf height can be adjusted for the more common sizes.

Instead, I believe the underlying reason for open stacks is the desire to provide a subject approach through classification. In the opinion of Elaine Svenonius, "the main use for classification, at least in the United States, has been to facilitate browsing of books on shelves.... Shelf browsing is an entrenched American tradition, but an activity about which relatively little is known."[3] Some research indicates a tendency for patrons to find materials through browsing, that is by using classification as a broad subject approach. In fact, many patrons in all types of libraries avoid the card catalog by immediately heading towards the parts of the collection where they are most likely to find their materials.[4] As further evidence, proponents of assigning classification numbers to everything explicitly state that they want to bring all like materials together intellectually no matter what their format. For example, the following is a quote from a report on the meeting of the Subject Analysis Committee Subcommittee on Subject Access to Microcomputer Software: "Finally, the fixation of classing all

software in Dewey 001 or LC QA must be avoided; instead the material must be distributed where it would logically fall if it were in book form, even if it needs to be gathered in a special collection.[5]

In a sense, the dependence upon classification proves that we mistrust the type of access that we provide through the card catalog. Yet, unlike the British tradition of a classed catalog, American libraries seldom make the shelf list available to the public but use it as a location and inventory device within cataloging areas. Thus, the patron may search on the shelves by classification order but is not allowed access to the bibliographic tool arranged in this manner.

2. Uniformity vs. diversity of classification schemes. Almost as a corollary of the first factor, American libraries take classification seriously. As I see it, one of the tensions has been the pull towards uniformity balanced against tendencies towards diversity. While uniformity has won out in the sense that most American general libraries use either the Dewey Decimal Classification or the Library of Congress Classification, I would like to look into the various factors involved in this struggle.

The United States has a long tradition of local independence which extends to libraries, and the extent of library autonomy would surprise many of our foreign guests. In principle, almost all libraries have the theoretical possibility of choosing whatever classification scheme they wish with any local modifications, or they can set up their own classification. Thus, no centralized body had a part in imposing the existing uniformity. Furthermore, as

many have commented upon, the Library of Congress is not a national library in the legal sense of the term but instead the Library of Congress. No federal legislation exists to force any library to accept LC's classification scheme. In a similar fashion, the Dewey Decimal Classification is supported by a self sustaining foundation which has no force except the force of reason to have any American library use D.D.C. Finally, we do not even have a centralized educational system for librarians which could help impose uniformity by educating all librarians under the same guidelines.

Yet what reasons have led American libraries of all sizes from the major academic research institutions of the world to small school and public libraries to the present degree of uniformity? Here I see many forces at work. Melvil Dewey was a towering figure during the latter part of the 19th century, and I do not underestimate the power of his missionary zeal in spreading his classification system. Yet, I find deeper reasons within the American character for subsequent developments. In addition to a common language,[6] the United States has a common cultural unity which is surprising given the multiplicity of political jurisdictions. The very American centered bias of our classification schemes helped to shut out any competitors from abroad.

On a practical level, pragmatism and a willingness to cooperate led to the general sentiment, though there have been notable exceptions,[7] that the same work should not be done over and over again in different libraries. In fact, the spread of the Library of Congress Classification can probably be attributed to LC's card

distribution service, which began in 1904. The maintenance of a separate classification scheme requires an intellectual effort of a high order which most libraries feel should be spent elsewhere, especially now that so much cataloging is done by library assistants who do not have a professional library degree.

I would speculate that the patron-centered philosophy of American libraries has also played its part. I do not perceive librarians in the United States as setting themselves up as gate keepers to library materials. Part of our problem in gaining professional recognition and respect from our public has been our zeal to impart our skills to them and to make the library as comprehensible as possible. Our tradition favors teaching our patrons how to use the library on their own and minimizes the interaction of librarians with skilled and, all too often, unskilled users. If each library were to have its own classification system, knowledge gained in one library would not transfer to another.[8]

3. <u>The theoretical vs. the pragmatic</u>. My final point is the radically opposed philosophical basis of the two major American classification systems. From its inception, D.D.C. attempted to give a theoretical ordering of all human knowledge while LC's goal was the classification and ordering of items within one particular library.

Each system has it advantages. D.D.C. makes it easier to go systematically from the general to the specific by moving to the right in the classification number, but LC may use its variety of expansion devices at any point in the schedule to open up space for

new divisions of knowledge. Dewey is a theoretically closed system which can expand only in detail by adding decimal places while Library of Congress, though less rigorous in its intellectual framework, remains hospitable to change.

By its pragmatic approach, LC also may readily admit its American bias by stating that as an American library its collection priorities are weighted towards its own country and that literary warrant justifies its decisions to emphasize the American perspective. Dewey, on the other hand, in practice reflects much of the same decision making since its founder needed to deal with the world as it existed in his day and from his point of view, but is in theory more open to criticism on the grounds that a theoretical system should have avoided Western, American, Christian bias.

SURVEY OF CLASSIFICATION SYSTEMS IN THE UNITED STATES

The two major classification schemes in the United States are the Dewey Decimal Classification and the Library of Congress Classification.[9] In addition, I would like to briefly comment on the Superintendent of Documents system since it functions as classification in the many libraries that have a separate United States government documents collection.

The Dewey Decimal Classification is probably the one most familiar to those of you in the audience since it has achieved the greatest acceptance outside of the United States. Even in this country, academic librarians in Library of Congress Classification

libraries probably encountered Dewey in library school and continue to use it if they have any contact with their local public library. Melvil Dewey published the first edition of the Dewey Decimal Classification in 1876, and the most current edition, the 19th, appeared in 1979 in three volumes. Responsibility for the D.D.C. rests with Forest Press, a wholly owned subsidiary of the Lake Placid Education Foundation which Melvil Dewey set up to carry on his work. Editorial work is carried out under contract at the Library of Congress though Forest Press continues to market and to publish D.D.C. One recent innovation is a decision to implement continuous revision of D.D.C. between editions so that changes do not occur all at one time when a new edition appears.

Dewey's basic structure divides all knowledge into ten groupings and then further subdivides each on a decimal basis such that each position to the right denotes a greater degree of specificity. Each classification number must have the three digits to the left of the decimal point while division to the right of the decimal is theoretically unlimited. While great specificity is possible, smaller libraries may shorten long numbers to meet the needs of their collection; and the Dewey Classification numbers distributed in Library of Congress bibliographic records indicate rational truncation points.

D.D.C. employs seven auxiliary tables for number expansion to allow for modification of the regular classification number by specific aspect. Currently, the four most commonly used tables are standard subdivisions, geographical subdivisions, individual literature subdivisions, and individual language subdivisions. One

current proposal by Arnold S. Wajenberg is to provide machine readable delimiters between the main part of the classification and the subdivisions in order to permit more sophisticated machine readable searching.[10]

Within the United States and Canada, 85% of all libraries use the Dewey Decimal Classification.[11] It is generally considered more suitable for public libraries and smaller academic institutions though several major academic research libraries, University of Illinois and Duke for example, remain committed to D.D.C.. One problem for any library which collects non-Western publications is the relatively little space given to topics of concern to Third World Countries in comparison with the space devoted to American and Western European thought.

The Library of Congress Classification, though perhaps more of interest to our foreign guests, defies any brief description. Never intended as a universal classification scheme, its purpose was to classify those books held by the Library of Congress; and revisions occur as needed. Influenced by Charles Ammi Cutter's Expansive Classification, J. C. M. Hanson, head of the Catalog Division, and Charles Martel, chief classifier, introduced the new classification scheme at the Library of Congress. Class Z, Bibliography, was the first published schedule and appeared in 1898. The Library of Congress Classification is not complete because law classification schedules for much of the world have not yet appeared.

To indicate the difficulties in describing the system, let me make one extensive quotation:

> The scheme consists of 28 different schedules, each having different authorship, possessing varying dates of origin and revision, separate indexes and individual auxiliary tables. No philosophic principle governs the scheme's constitution and no manual of instruction for use exists for the scheme as a whole. It was intended as an "in-house" system for catalogers classifying materials for the United States Congress rather than a broadly conceived universal scheme. The system does not necessarily suit libraries other than the Library of Congress nor was its purpose ever to do so.[12]

Be that as it may, I will nonetheless try to describe briefly some of the more general features of the scheme since it is used in about 15% of American libraries including "virtually all university libraries and half of all college and junior college libraries."[13] The first line of each classification number is made up of an alphabetical portion of from one to three letters and a number portion. As with the Dewey Decimal Classification, numbers may be divided decimally; but this does not indicate a clear hierarchical arrangement since simple call numbers in parts of the schedule may indicate a high degree of specificity. If further subdivision is required, a cutter number may be given on the second line, usually using the topic of the subdivision to form the cutter such as .B3 for Bavaria or .N4 for nickel. Again, numbers are expanded decimally as needed to make any new topic fit in alphabetically with the others in the list. Especially in Class H

(Social Sciences), complex auxiliary geographic tables exist which are needed to formulate the call number for specific areas of the world. While such classification numbers logically divide the world according to geographic areas, they are not added on at the end as in the Dewey Decimal Classification nor would they be accessible as country indicators in a machine readable environment.

Why then is the Library of Congress Classification used by so many academic research libraries within the United States? First, there is the pragmatic reason indicated above that Library of Congress cataloging data provides the Library of Congress call number including the book number. Second, the system is so immense and hospitable to change that new topics can be added with a minimum of disruption. Third, most American academic research libraries have reasonably similar collecting policies to those of the Library of Congress though on a smaller scale. Except for items of purely local interest which can be accommodated easily in Class F (Local American History, F1-975), Library of Congress purchases a high percentage of what American research libraries buy and groups the materials in a way understandable to American scholars. Fourth, unlike Dewey, the call numbers remain short even though the classification provides a high degree of specificity. In fact, LC's flaws may become apparent only as future attempts at developing call number searching reveal its lack of underlying structure.

While perhaps not thought of as a classification scheme, Superintendent of Documents numbers function as such in many American libraries. For practical reasons, many American libraries do not individually catalog United States government documents but

instead shelve them by the Superintendent of Documents number. Since government bibliographic aids such as The Monthly Catalog give this number, patrons can use it to retrieve desired items. In a sense, this type of access to government documents is similar to the use of indexes for periodicals.

Superintendent of Document numbers are based upon the organizational structure of the United State government. The first part of the classification number is a letter, often a mnemonic for the issuing agency, then a number with numerical divisions as necessary to place the issuing agency within the hierarchical structure, and finally a book number. I find this system of interest since it goes contrary to the principles indicated above about the need for subject access through classification, and indeed some American librarians would like to integrate government documents into the general collection.[14] Yet even here a subject grouping exists since the government is arranged by subject interest divisions. Most information on education emanates from the Department of Education while items on national defense will come from the Department of Defense. While results are not conclusive, preliminary research indicates that American library users expect to be able to browse successfully with the Superintendent of Documents classification.[15]

CURRENT ISSUES IN CLASSIFICATION

I believe that classification currently receives much less attention both on the theoretical and practical levels than verbal subject access though I almost hope that some of you will disagree

with this perception during the question and answer period. I see several reasons for this lack of attention. The first is the very practical one that while American libraries use two different classifications, almost all use Library of Congress Subject Headings or Sears Subject Headings, which share the same philosophy and differ on most points only in comprehensiveness.

Second, since subject headings are intended to allow library users to find materials in our bibliographic tools, they must of necessity match the language of the users. Yet as we all know, language is in a constant state of change, especially in this era of rapid technological innovation. As new concepts appear, as old ones die, as terminology is modified, subject headings, to be easily used, should reflect these linguistic changes. However, thesaurus and subject heading list revision always lags behind in the same way as no dictionary is ever completely up to date. Since the gap between library subject headings and current usage is easy to see, it occasions comment both by librarians and patrons.[16]

Classification, on the other hand, is already a rational construct and further removed from the day-to-day life of our patrons. I believe that for most of them call numbers already reside in the land of technical librarianship. Users were much more likely to tell us that "European War, 1914-1918" was a silly term for World War I than that that 450.7 or HG459 is a bad call number.

The third reason follows from the second. American librarians hope for a solution to subject heading problems from computers while they are at present less likely to see any radical change in

classification as a result of the growth of online catalogs. While I realize that this is not a totally accurate statement since we have projects such as the current effort to examine the role of the Dewey Decimal Classification in the online catalog,[17] Americans are more excited about the possibility of radically improved verbal subject access from key word searching, Boolean logic, customized authority control, and the like.

Finally, and I will dwell more on this later, librarians are more conservative about classification than about subject headings because they do not want to remark the books. While various strategies may be used to provide for change in subject headings such as explanatory cross references, updating the cards or machine readable records, etc., no easy solutions exist for implementing changes in call numbers since the previously marked books exist as physical objects requiring expensive human labor even in this computer era.

One issue from the recent past that is no longer of much concern is the debate about the relative suitability of the Dewey Decimal and Library of Congress Classifications for use in large academic research libraries. As a side effect of the rapid growth of academic research libraries during the 1960's, many changed over to the Library of Congress Classification, some from their home grown systems but most from Dewey.[18] As libraries expanded rapidly, the larger collections provided a rationale for LC since it was felt that Library of Congress Classification provided finer distinctions for large collections and more nearly met the needs of academic researchers. Furthermore, Library of Congress printed cards and

machine readable records do not include Dewey numbers for many foreign language and esoteric research materials; and the expanding libraries found that they could get a much higher percentage of ready-to-use classification numbers if they adopted the Library of Congress Classification.

This trend diminished in the 1970's and early 1980's for various reasons. Many of the libraries for which the change was suitable had already done so. In addition, the last ten years have been a time of retrenchment. Libraries would have more difficulty in obtaining the resources to make the transition. Furthermore, our patrons object to having items shelved in different parts of the library with different classification schemes. In fact, in a recent survey of library information seeking behavior at my own institution, faculty volunteered in the comment section that they would like to see the two parts of the collection brought together.[19] When the decision to change classification was made in former richer times, some hope was usually given for the reclassification of the older materials into the new classification. Today, when libraries feel themselves fortunate to have the staff to cope with current receipts, reclassification no longer is a prudent use of resources. Thus, it does not seem likely that many academic library users of Dewey will switch to the Library of Congress Classification in the near future.

The main philosophical issue for the Dewey Decimal Classification is the appropriateness of Phoenix schedules. As indicated earlier, D.D.C. is a closed system whose overall structure is relatively inflexible. Forest Press and its advisory boards are

faced with difficult choices. Without revision, D.D.C. will deviate more and more from the current structure of knowledge as ways of thinking undergo substantial change. Even when the numeration is appropriate, any area with a rapid expansion of information may require longer and longer numbers to accurately reflect new developments. To resolve this difficulty, D.D.C. has decided to target a limited number of major revisions for each new edition. These revisions, called the Phoenix schedules, start off from a new philosophical base to develop a revised schedule which often reuses an existing call number range. There is an obvious difficulty here for libraries. They already have items under the old numbers which they must either change, or they must live with the intermixture of two varying meanings for similar Dewey numbers. In a sense, the conflict is unresolvable because without mechanisms for change D.D.C. would eventually become obsolete as it no longer reflected the world of knowledge of its time. On the other hand, as stated earlier, reclassification costs money and staff time that libraries can ill afford. I would submit, in fact, that there would be more opposition to this policy if it were not for the fact that many small to medium public libraries expect to withdraw obsolete books over a period of time and that this withdrawal policy serves as a self-correcting device for intermingled classification problems.

The Library of Congress Classification handles the problem of substantive revision by creating whole new areas and declaring the old numbers obsolete. Not being bound by a decimal arrangement, L.C.C. can add letters to the alpha portion and expand to infinity if necessary in the numerics. Examples of such changes in recent years are the creation of BQ in the religion section for Buddhism

and the expansion of Polish history by multiplying each number by 10. Thus, DK401-DK443.7 became DK4010-DK4800. While reclassification would still be necessary to bring all the materials together, at least a dual meaning for the same number does not occur.

Both systems, however, still reflect the ordering of knowledge that prevailed during the period of their birth. Psychology (BF in the Library of Congress Classification and 150-159 in Dewey) remains a part of Philosophy though literary warrant today would give it much greater prominence in a new classification scheme.

I do not want to spend time on the role of classification in the online catalog. This issue is not limited to the United States but will have world-wide implications. I am especially interested in reading the results of Karen Markey's project on subject access through D.D.C.[20] In addition, I would suggest consulting "Use of Classification in Online Retrieval" by Elaine Svenious for a discussion of the theoretical possibilities as she sees them.[21]

INTERNATIONAL IMPLICATIONS OF CLASSIFICATION IN THE UNITED STATES

The United States, as stated previously, has had world wide influence in the area of library classification. Even the Bliss Classification, not widely used in the United States, has had its adherents in Great Britain.[22] Forest Press, as keeper of the Dewey Decimal Classification, takes its international responsibilities seriously and supports translations and adaptations of D.D.C. into foreign languages. The relationship between Forest Press and the issuing body varies from country to country, but

currently Dewey or some subset is available in at least a dozen languages.[23] On the other hand, the Library of Congress Classification remains in principle a classification scheme for its own collection which others use at their own risk. While Canadiana, the Canadian national bibliography, provides Library of Congress classification numbers with an expansion in the F schedule for Canadian history (FC), the Library of Congress makes no claim for the international aspects of its system. While I am aware that L.C.C. is used internationally at least in the United Kingdom, Canada, and Australia, I was unable to find a comprehensive list of L.C.C. foreign libraries. Perhaps Forest Press is more receptive to an international marketing effort because its existence depends upon the continued acceptance of the Dewey Decimal Classification while the Library of Congress is assured of its continued existence through the support of the American taxpayer.

As I stated in my paper last year at the Nairobi IFLA Conference in another context,[24] I am convinced of the advantage of classification as a subject retrieval device across international and linguistic boundaries in comparison with verbal subject headings. Classification is tied to the concept rather than to the naming of the concept and as long as the concept in the real world remains the same in each language all that needs translation are the descriptive and index terms in the classification schedule. In a certain sense, classification can serve as a scientific name for unambiguous concepts in the same way as Latin serves as the basis for scientific names in biology. As Elaine Svenonius states, "a classification such as DDC that has already been translated into many languages comes ready-made as a switching mechanism. With the

use of such a classification, a user's search terms entered into the system in one language can be switched through DDC numbers to retrieve documents in several different languages."[25] Where the system breaks down is in the area of human constructs since political, economic, and educational systems vary from country to country even where the language is the same. These problems, however, are minor in comparison to the difficulties of transferring a subject heading structure with all its linguistic differences from one country to another, even when they both share a common language.[26] Furthermore, classification can adapt to changing and varying terminology as long as the underlying concept remains the same.

Classification also has the advantage of bringing together like materials in an efficient way, especially in small collections. While the user of a small collection may have difficulty in following the subject heading reference structure to broader, narrower, and related terms when the complexity has been established for large collections, classification can group together within a restricted number of records all the items on a subject, at least as long as the subject is the one that determined the classification number of the book.

Thus, the classification numbers provided by United States cataloging agencies, both in the Dewey Decimal and in the Library of Congress Classifications, may be more useful to the international community than the subject headings in the same record. Perhaps our Section as part of its international charge should suggest and encourage research on the whole issue of classification within the

international context. I for my part would like to see more research done within the American library community to see how the Dewey Decimal and the Library of Congress Classification schemes can serve all libraries.

THE FUTURE OF CLASSIFICATION IN THE UNITED STATES

I am confident that classification will continue to play a substantial role in United States librarianship. For the reasons given above, the innate conservatism of librarians who do not wish to cope with remarking existing items will work against substantive modifications in existing classification schemes. Furthermore, except for special libraries, I do not foresee the invention of any new schemes in the near future. If anything, it is the online catalog that will have an impact in the medium to long term future. If I can hazard any guess about the future, I would suggest that the computer may bring back the classed catalog where classification once more takes on a more important role in subject access. One classification number will indeed be the call number and provide information as to the physical location of the book, but subsidiary numbers for subsidiary subject aspects will provide an alternative access point in the online record as the current shelf list becomes a subject access tool. Just as the library patron today expects to find all relevant items under the appropriate subject heading, the library patron of the future may expect the same assurance as he or she browses through the online shelf list. Though knowledgeable in the intricacies of the card catalog, I have often used browsing as an access tool when I wanted several books on a subject without needing specific titles. An online catalog which would provide the

same capability on the computer screen while also bringing to my attention both other related classification numbers and items in other locations would be a definite improvement.

I would like to thank the Section for the opportunity to speak on classification in the United States today. I hope that I have not wandered too far afield and have not misjudged the interests of my audience by spending more time on the issues than on the technical details of American classification.

NOTES

[1]I would like to acknowledge the assistance of Elaine Broadbent, Head of Original Cataloging, Marriott Library, in the preparation of this paper.

[2]The Bowker Annual of Library & Book Trade Information, 29th ed. (New York, 1984), p. 347.

[3]Elaine Svenonius, "Directions for Research in Indexing, Classification, and Cataloging, Library Resources and Technical Services, 25 (January/March 1981), 94-95.

[4]Other recent articles on browsing are:

R. J. Green, "The Effectiveness of Browsing," College & Research Libraries, 38 (July, 1977), 313-316.

Patricia Willard and Viva Teece, "The Browser and the Library," Public Library Quarterly, 4 (Spring, 1983), 55-63.

Johanna Ross, "Observations of Browsing Behavior in an Academic Library," College & Research Libraries, 44 (July, 1983), 269-276.

[5]Gregory R. New and Winton E. Matthews, "CCS/SAC Subcommittee on Subject Access to Microcomputer Software," Library of Congress Information Bulletin, 43 (September 27, 1984), 306.

For another quote on nonprint materials, we have: "General classifications have largely ignored nonprint materials.... By

assigning detailed classification numbers to everything and making those numbers independent of the way objects are stored, we would be treating nonprint materials fairly and making them equally accessible to the library user." Michael Gorman, "The Longer the Number, the Smaller the Spine," American Libraries, 12 (September, 1981), p. 498.

[6]It will be interesting to see what changes occur as Spanish becomes more important as a second language within the United States and as library service increases for the Spanish-speaking population.

[7]There have been some exceptions to this uniformity. Until about 1970, the Yale University Libraries used their own classification system before switching to Library of Congress Classification, but even here much of the classification merely transferred LC classification numbers to a new notation scheme.

[8]Americans have believed in standardization since the introduction of interchangeable parts by Eli Whitney. I remember reading an essay which contended the the similar arrangement of our fast food restaurants and convenience grocery stores all across the United States contributed strongly to our sense of national unity.

[9]For this brief presentation of the Dewey Decimal and Library of Congress Classifications, I have borrowed heavily from the summary in:

Mildred Marlow Deeming, Introduction to Cataloging and Classification, 9th ed. (Jefferson, N.C., 1981).

[10]Arnold S. Wajenberg, "MARC Coding of DDC for Subject Retrieval," Information Technology and Libraries, 2 (September, 1983), 246-251.

[11]John P. Comaromi, "Use of the Dewey Decimal Classification in the United States and Canada," Library Resources and Technical Services, 22 (Fall, 1978), 403.

[12]Ibid.

[13]Downing, p. 97.

[14]For a discussion of this tension between integrating government documents and maintaining them in special areas, see:

Peter Hernon and Charles McClure, Public Access to Government Information : Issues, Trends, and Strategies (Norwood, N.J., 1984), p. 218.

[15]"Since browsing patterns for government publications have not been reported in library literature, it may be interesting to describe those of four social scientists who listed this as a means for locating publications...." Peter Hernon, Use of Government Publications by Social Scientists (Norwood, N.J., 1979), p. 56.

See also:

Hernon and McClure, Public Access to Government Information: Issues, Trends, and Strategies, p. 218.

[16]The literature on the problems of Library of Congress Subject Headings is immense. Sanford Berman is perhaps the major proponent of changing LCSH to more closely match contemporary American usage.

[17]Karen Markey, "Subject-Searching Experiences and Needs of Online Catalog Users: Implications for Library Classification," Library Resources and Technical Services, 29 (January/March, 1985), 34-51.

[18]Robert L. Mowery, "The 'Trend to LC' in College and University Libraries," Library Resources and Technical Services, 19 (Fall, 1975), 389-397.

Also on the same topic: Comaromi, p. 403.

[19]Elaine Broadbent, "A study of Humanities Faculty Library Information Seeking Behavior," Unpublished manuscript submitted to the Cataloging & Classification Quarterly, p. 13.

[20]For details of this project, consult the article as cited in Note 16.

[21]Elaine Svenonius, "Use of Classification in Online Retrieval," Library Resources and Technical Services, 27 (January/March, 1983), 76-80.

[22]"To our knowledge the Bliss Classification is used completely as a system in only one library in the U.S.,--that of City College where it was instituted by its author. The system is more widely used and apparently more popular in Great Britain." Maurice F. Tauber and Edith Wise, Classification Systems (New Brunswick, N.J., 1961), p. 165.

[23]Forest Press has informed me that authorized editions of the Dewey Decimal Classification are available in French and Spanish and are in preparation for Arabic and Italian. Local editions have been developed in Afrikaans, Hebrew, Indonesian, Korean, Malayan, Thai, Turkish, and Vietnamese.

[24]Robert P. Holley, The Consequences of New Technologies in Classification and Subject Cataloging in Third World Counties: The Technological Gap (Nairobi, 1984). IFLA Document 32-CLAS-1-E.

[25]Svenonius, p. 80.

[26]For the problems of adapting Library of Congress Subject Headings within an Australian context, see:

John McKinlay, "Australia, LCSH and FLASH," Library Resources and Technical Services, 26 (April/June, 1982), 100-108.

IFLA General Conference: Chicago 1985

Division: Bibliographic Control

Section: Classification and Subject Cataloging I

48-CLAS-2-E

Dewey Decimal Classification Online Project: Integration of a Library Schedule and Index into the Subject Searching Capabilities of an Online Catalog

by

Karen Markey

Research Scientist

OCLC Online Computer Library Center

Dublin, Ohio 43017 USA

and

Anh Demeyer

Systems Analyst

OCLC Online Computer Library Center

Dublin, Ohio 43017 USA

For internal use only:

To be brought to meeting no: 87

Remarks:

to be presented orally ☒

for distribution only ☐

Classification and Subject Cataloging I

K. Markey and A. Demeyer (USA)

Dewey Decimal Classification Online Project

Abstract

The Dewey Decimal Classification (DDC) Online Project is supported by the Council on Library Resources, OCLC, and Forest Press (publishers of the DDC). In the research project, subject-rich terms from the DDC Schedules and Relative Index are incorporated into an online catalog as searchers' tools for subject access, browsing, and display. The machine-readable DDC used in this project was converted from the print tapes that originally served to produce the nineteenth edition of the printed DDC using computerized photocomposition.

Four features of the DDC are employed to help searchers browse for and match their own subject terms with the online catalog's terminology: (1) subject terms in the DDC Schedules (e.g., headings, scope notes, classhere notes, including notes), (2) hierarchical arrays of related terms in the DDC Schedules, (3) subject terms in the DDC Relative Index, and (4) alphabetical displays of DDC Relative Index entries. Using the expanded subject information from these sources, users of the experimental online catalog have a greatr chance of finding exactly the topic they are seeking. The effectivenessof the DDC in an online catalog was tested in online retrieval experiments withlibrary patrons and staff at the Library of Congress, the New York State Libray, the Public Library of Columbus and Franklin County (Ohio), and the Mathematics Library of the University of Illinois at Urbana-Champaign. Findings willbe available in October 1985.

Classification in Online Catalogs

The Dewey Decimal Classification (DDC) Online Project, which is supported by the Council on Library Resources, OCLC, and Forest Press, focuses on how the DDC can be used as an online searcher's tool for subject access, browsing, and display in an online catalog. In this research project, the Relative Index and Schedules of the DDC have been integrated into the subject searching features of an experimental online catalog. The experimental catalog is being evaluated by library patrons and staff in four libraries: the Social Science Reading Room of the Library of Congress, the New York State Library, the Public Library of Columbus and Franklin County (Ohio), and the Mathematics Library of the University of Illinois at Urbana-Champaign.

Library classification offers a logical approach to subject searching and increases the amount of subject information included in bibliographic record displays. Library classification can also make possible search strategies that are not presently supported by online catalogs.

In American libraries, classification schedules and indexes have chiefly been the tools of catalogers and classifiers used behind the closed doors of technical processing departments. Classification has been visible to library users only in the shelf arrangement of library materials. Patrons typically find an item of interest by searching for subject headings or subject and title keywords in the online catalog, recording a call number from the catalog, locating the book bearing that number on the shelf, and then browsing the bookshelves for other similar books. Here we will show how the index and schedules of a library classification can be incorporated into an online catalog to enhance subject access to its contents, help searchers browse for other terms related to their search terms, and provide additional content information in bibliographic records to help online searchers assess the relevance of retrieved items.

Rationale for the Project

Classification schedules were first mechanized for bibliographic retrieval by Freeman and Atherton (1968), who demonstrated that Universal Decimal Classification (UDC) could be used as an entry vocabulary for online subject searches. At the time of their study, few online retrieval systems were operational and even fewer (if any) of these were available to end users such as library patrons and reference librarians.

Thirteen years after the UDC study, in the fall of 1981, the OCLC Office of Research developed a proposal to examine how a library classification could be used to assist library catalog users in performing online subject searches. This proposal emerged from ongoing research at OCLC. Several developments shaped our ideas on the potential of DDC in the environment of an online public access catalog:

1. A nationwide study of the online catalog sponsored by the Council on Library Resources showed that subject searching was the predominant approach to accessing online catalogs and that users experienced many difficulties when they performed subject searching in existing online catalogs (McCarn 1983).
2. The significant growth in the design and implementation of online catalogs was recognized by library administrators involved in online catalog planning and development, who took part in a conference sponsored by the Council on Library Resources (CLR) (Kaske and Ferguson 1980).
3. A report on subject access (Mandel 1981, 16) prepared for the Council on Library Resources, recommended that terms from a classification schedule be appended to bibliographic records to enhance subject

access; and a CLR-sponsored invitational meeting on subject access recommended that a library classification in machine-readable form be designed and distributed (Russell 1982, 70).

4. The Visiting Distinguished Scholar at OCLC in 1981, Pauline Cochrane, had conducted related research on the UDC in an online bibliographic retrieval system (Freeman and Atherton 1968) and on improved subject access to bibliographic records (Atherton 1978) which influenced the Office of Research.

The nineteenth edition of Dewey Decimal Classification (DDC) had been produced by computerized photocomposition, and the Office of Research requested and received permission from the publisher, Forest Press, to use this machine-readable data for research purposes (Custer 1979, 1:xii). In January 1984, with the support of CLR and Forest Press, the OCLC Office of Research embarked on a study of DDC as a user's tool for subject access, browsing, and display. The machine-readable DDC used in the Dewey Decimal Classification Online Project was converted from print tapes used originally to produce the nineteenth edition.

The project will answer the following three research questions:

1. Is the machine-readable DDC suitable to implement in an online catalog as a searcher's tool for subject access, browsing, and display?
2. When used as a searcher's tool for subject access, browsing, and display, does the online DDC improve the search performance of subject searchers at the online catalog?
3. Do subject searchers prefer an online catalog in which the DDC is available as a searcher's tool to an online catalog without online DDC?

The first question is being addressed by representatives of the four participating libraries, project consultants, staff in the Dewey Decimal Classification Division at the Library of Congress, Forest Press staff, and the DDC Online Project team. To answer the second and third questions, online search experiments are being conducted with library patrons and staff. The retrieval experiments used in this research project are based on the experiments used in a comparative evaluation of two prototype online catalogs at the National Library of Medicine (Siegel et al. 1983).

The Experimental Online Catalogs

Two experimental catalogs have been developed for the research project. The first catalog, called SOC (for Subject Online Catalog), has the traditional subject searching capabilities of online catalogs, namely, by subject heading and subject and title keywords. The second catalog, called DOC (for Dewey Online Catalog), has both traditional subject searching capabilities and enhanced subject searching capabilities based on the DDC Relative Index and DDC Schedules and notes. Bibliographic records in selected subject areas were obtained from each of the four participating libraries. Separately, these four bibliographic record collections have been indexed into both SOC and DOC experimental online catalogs.

Both SOC and DOC are easy to search. The catalogs are menu-based systems and always provide a list of options at the end of every system response which tell the user the possible next steps in the search. Both systems also provide an initial screen that describes the available subject search options and the system responses. Reading this initial screen is mandatory, but further instruction in the use of the catalogs is unnecessary.

The Subject Online Catalog has the traditional subject searching capabilities of online catalogs, namely, by subject heading and subject and title keywords. Subject searching options in SOC are described in Table 1.

TABLE 1. Subject searching in the Subject Online Catalog (SOC)

SOC Option Code	Functional Description	Source of Subject Terms for Indexing	Wisest Subsequent User Option(s)
SA	Search for subject and browse an alphabetical list of subjects near your subject	Assigned subject headings from the library's bibliographic records	Select a listed subject heading to display retrieved items Browse backward or forward in list of subject headings
SD	Search directly for subject and retrieve number of items on your subject	Keyword (implicit Boolean "and") search of title, subject heading, series, and note words in the bibliographic records	Display retrieved items
SC	Search directly for call number and retrieve number of items with your call number	Class numbers in the library's bibliographic records	Display retrieved items

A detailed description of SOC capabilities is given in an interim report of the project (Markey and Demeyer 1985, 6-13). The major difference between the Subject Online Catalog and Dewey Online Catalog is that the former's subject searching capabilities have not been enhanced by the DDC Relative Index or Schedules. SOC allows class number searching, but indexed class numbers are obtained from the library's bibliographic records.

In the Dewey Online Catalog the DDC Relative Index and DDC Schedules and notes have been integrated into DOC subject searching capabilities. Subject searching options in DOC are described in Table 2.

TABLE 2. Subject searching in the Dewey Online Catalog (DOC)

DOC Option Code	Functional Description	Source of Subject Terms for Indexing	Wisest Subsequent User Option(s)
SA	Search for subject and browse an alphabetical list of subjects near your subject	DDC Relative Index entries that match numbers in the library's bibliographic records	Select a listed Relative Index entry to display item records Browse forward or backward in the alphabetical list of Relative Index entries Select a listed Relative Index entry to obtain a subject outline from the DDC Schedules
SS	Search for subject and browse an outline of subjects related to your subject	Keyword (implicit Boolean "and") search of DDC Relative Index entries, DDC Schedules and note words, and the first subject heading listed in the library's bibliographic records which direct users to outlines of DDC Schedule captions in the classification area(s) that have records bearing user-entered terms	Select a listed DDC Schedule caption to display retrieved items Browse more specific DDC Schedule captions
SD	Search directly for subject and retrieve number of items on your subject	Keyword (implicit Boolean "and") search of title, subject heading, series, notes, DDC Relative Index entry, and DDC Schedule and note words in bibliographic records	Display retrieved items

TABLE 2. (continued)

DOC Option Code	Functional Description	Source of Subject Terms for Indexing	Wisest Subsequent User Option(s)
SC	Search for call number and browse an outline of meanings of call numbers	Class numbers in the library's bibliographic records, which directs users to outlines of DDC Schedule captions in the classification area matching user-entered class number	Select a listed DDC Schedule caption to display retrieved items. Browse more specific DDC captions

Subject Searching with DOC

Initially, the searcher in the Dewey Online Catalog is provided with a menu containing the following four subject searching options:

SA Search alphabetically for subject

SS Search subject outline for subject

SD Search directly for subject

SC Search for call number

A DOC user who performs a search alphabetically (SA) obtains a list of Relative Index entries (Figure 1). The user can then browse forward or backward in this list or display bibliographic records whose class number matches the class number corresponding to the Relative Index entry. Just as entries in the printed DDC Relative Index direct users to a specific area in the DDC Schedules, the selection of a Relative Index entry in DOC directs the online user to captions and notes in the online DDC Schedules.

DOC provides users with an option to search subject outlines (SS) composed of captions from the DDC Schedules. DOC performs a keyword and implicit Boolean search of the first subject heading listed in the subject heading fields of the bibliographic record, and terms from the DDC Schedules and Relative Index. DOC directs the user to a display of headings from the pertinent area(s) of the DDC

```
What would you like to do? > sa multivariate analysis

>>>>> YOUR REQUEST ... sa:multivariate analysis

CHECK THIS LIST OF SUBJECTS IN ALPHABETICAL ORDER.  IF YOUR SUBJECT IS NOT
LISTED TRY ANOTHER SEARCH OPTION.

LINE   ITEMS TOPICS--ASPECTS
   A       1 Multiplication--arithmetic
   B      14 Multistage programming--mathematics
   C      13 Multivariate analysis--statistical mathmetics
   D      15 n
   E      15 n-ary invariants algebra
   F      66 Nets (mathematics)
   G       6 Nets (mathematics)--combinatorial topology
   H       7 Neumann function
   I      60 Nilpotent algebras and groups
   J     161 Non-Euclidean geometries

OPTION DESCRIPTION
    BF   Browse forward
    BB   Browse backward
    ST   Search subject outline for a topic from the list
    SS   Search subject outline for "multivariate analysis"
    SD   Search directly for "multivariate analysis"
    SL   Select line(s) to display items
    SO   Start over
```

FIGURE 1. Relative index entries and options displayed in a DOC alphabetical search (SA)

Schedules (Figure 2). Class numbers and captions from the DDC Schedules are listed. A user can select any one of the topics displayed to browse a list of more general or more specific topics. The user can also view notes from the DDC Schedules to get more subject information about a listed topic. This information may be in the form of an example note, including note, or class here note, or it may be a general note, such as a scope note. DOC users can also display bibliographic records in shelflist order. Postings, which are cumulative, indicate the number of books assigned the exact number listed and the number of books assigned a class number beginning with the number listed.

Users can also search directly for a subject by choosing the SD option at the beginning of a search and entering a keyword or phrase (Figure 3). DOC

```
What would you like to do? > ss multivariate analysis

>>>>> YOUR REQUEST ... ss:multivariate analysis

Please wait ....
YOUR SUBJECT IS FOUND UNDER 519.53 IN THIS OUTLINE OF RELATED SUBJECTS.

GENERAL TOPIC : 519.5 Statistical mathematics.
   LINE     ITEMS     LOCATION     TOPIC
      A         3      519.52      Theory of sampling.
      B        53      519.53      Descriptive statistics and multivariate
      C        28      519.54      Statistical inference.
      D        12      519.55      Time series analysis.
      E         7      519.56      Statistical hypothesis testing.

OPTION      DESCRIPTION
    BG      Browse more general topics
    BS      Browse more specific topics
    GN      Get detailed notes for a topic
    SL      Select line(s) to display items
    RP      Return to perspectives list
    SO      Start over
```

FIGURE 2. DDC Schedule captions and options displayed in a DOC subject outline search (SS)

tells the user the number of items retrieved containing each term in the user-entered phrase and the number of items containing all terms entered. DOC performs a keyword search of title, subject, selected notes, and series fields, and unlike SOC, it also searches fields added to the bibliographic record from the DDC Schedules and Relative Index.

Searching by call number (SC) is the fourth subject searching option for DOC users (Figure 4). Unlike SOC, which merely reports the number of books assigned the exact class number and the number of books assigned a class number beginning with the number listed, DOC responds with (1) a list of topics (or headings) from the DDC Schedules at the same level in the hierarchy as the entered class number, and (2) a list of options that allow the user to browse more general or specific headings, get more information about a listed heading, or display bibliographic records assigned to the heading.

```
What would you like to do? > sd
Enter subject to search > multivariate analysis

>>>>> YOUR REQUEST ... sd:multivariate analysis
        62 ITEMS FOR "multivariate"
       743 ITEMS FOR "analysis"
        59 ITEMS IN SET 1 (multivariate analysis)

THE FOLLOWING IS A SUMMARY OF YOUR SEARCHES:

SET    ITEMS          REQUEST
  1      59           sd:multivariate analysis

OPTION DESCRIPTION
       DI     Display items
       SO     Start over
```

FIGURE 3. Number of retrieved items and options displayed in a DOC direct search (SD)

```
What would you like to do? > sc
Enter call number (between 510 and 519.86) to search > 519.5

>>>>> YOUR REQUEST ... sc:519.5

ITEMS ABOUT YOUR SUBJECT ARE TREATED UNDER 519.5 IN THE FOLLOWING OUTLINE.

GENERAL TOPIC : 519 Probabilities and applied mathematics.
LINE   ITEMS      LOCATION        TOPIC
   A     142      519.2           Probabilities.
   B      32      519.3           Game theory.
   C      74      519.4           Applied numerical analysis.
   D     241      519.5           Statistical mathematics.
   E      92      519.7           Programming.
   F      23      519.8           Special topics.

OPTION DESCRIPTION
    BG  Browse more general topics
    BS  Browse more specific topics
    GN  Get detailed notes for a topic
    SL  Select line(s) to display items
    SO  Start over
```

FIGURE 4. DDC Schedule captions and options displayed in a DOC call number search (SC)

Impact of the DDC Online Project

The implementation of Dewey Online Catalog developed in this research project is the first use of a library classification as a tool for subject access, browsing, and display in a library information system accessed by end users. From December 1984 to July 1985, the DDC Online Project team conducted retrieval tests with library patrons and staff at the four participating libraries, and the data will be analyzed in August and September 1985. The findings of the project will be presented in a final report to the Council on Library Resources in October 1985. At the project's conclusion, CLR will sponsor an invitational conference on online classification to disseminate the findings of the DDC Online Project and assess the project's impact on the availability and usefulness of machine-readable classification in online library catalogs.

The DDC Online Project will produce data and analyses that will contribute significantly to our knowledge of how library patrons and staff use online catalogs. The results of the research will be made widely available, especially to the library and information science community. It is our hope that the study will contribute to the enhanced design of library classification schedules and online catalogs and will extend patron use of library classifications in general.

REFERENCES

Atherton, Pauline A. 1978. Books are for use. Final report to the Council on Library Resources. Syracuse, N.Y.: Syracuse University. ED 156 131.

Custer, Benjamin A., ed. 1979. Dewey decimal classification and relative index. Vol. 1, Tables. 19th ed. Albany, N.Y.: Forest Press.

Freeman, Robert R., and Pauline A. Atherton. 1968. "File organization and search strategy using DDC in mechanized reference retrieval systems." In Mechanized information storage, retrieval and dissemination. Proceedings of the FID/IFIP Joint Conference, Rome, 14-17 June 1967. Amsterdam: North Holland.

Kaske, Neal K., and Douglas Ferguson. 1980 (September). On-line public access to library bibliographic data bases: Developments, issues, priorities. Final report to the Council on Library Resources. Columbus, Oh., and Stanford, Calif: OCLC and Research Libraries Group. ED 195 275.

McCarn, Davis B., ed. 1983. Online catalogs: Requirements, characteristics and costs. Report of a conference sponsored by the Council on Library Resources at the Aspen Institute, Wye Plantation, Queenstown, Md., 14-16 December 1982. Washington, D.C.: Council on Library Resources.

Mandel, Carol A. 1981. Subject access in the library catalog. Washington, D.C.: Council on Library Resources. ED 212 286.

Markey, Karen, and Anh Demeyer. 1985 (February). DDC online project. Interim report to the Council on Library Resources. Dublin, Oh.: OCLC.

Russell, Keith W. 1982. Subject access. Report of a meeting sponsored by the Council on Library Resources. Dublin, Oh., 7-9 June 1982. Washington, D.C.: Council on Library Resources. ED 230 205.

Siegel, Elliot R., et al. 1983. "Research strategy and methods used in conducting a comparative evaluation of two prototype online catalog systems." In National online meeting proceedings--1983, New York, 12-14 April 1983. Medford, N.J.: Learned Information.

DDC ONLINE PROJECT BIBLIOGRAPHY

Cochrane, Pauline A., and Karen Markey. 1985. Preparing for the use of classification in online cataloging systems and in online catalogs. Information Technology and Libraries 4, 2 (June). In press.

Markey, Karen. 1984 (April). DDC online project. Interim report to the Council on Library Resources. Dublin, Oh.: OCLC. Available from ERIC.

------. 1984. "The Dewey decimal classification as a library user's tool in an online catalog." In Challenges to an information society. Proceedings of the 47th annual ASIS meeting, 21-25 October 1984, Philadelphia, Pa. White Plains, N.Y.: Knowledge Industry Publications.

------. 1985. Subject-searching experiences and needs of online catalog users: Implications for library classification. Library Resources & Technical Services 29, 1 (January-March): 34-51.

Markey, Karen, and Anh Demeyer. 1984 (September). DDC online project. Interim report to the Council on Library Resources. Dublin, Oh.: OCLC. Available from ERIC.

------. 1985 (February). DDC online project. Interim report to the Council on Library Resources. Dublin, Oh.: OCLC. Available from ERIC.

IFLA
1985

IFLA GENERAL CONFERENCE, CHICAGO

Division: CONTROLE BIBLIOGRAPHIQUE

Section: CLASSIFICATION et CATALOGAGE - MATIERE

49-CLAS-3-E

SUBJECT INDEXING in FRANCE
TOOLS and PROJECTS

by

Suzanne JOUGUELET

Conservateur

Bibliothèque Municipale de LYON

France

KEYWORD(S):

Subject indexing, standardisation, France

Subject authority list, France

For internal use only:

To be brought to meeting no:

153

Remarks:

to be presented orally [X]

for distribution only [ ]

SUBJECT INDEXING in FRANCE : TOOLS and PROJECTS

At the beginning of the eighties, subject indexing in french libraries - except the most important institutions - was practised in a rather " individualistic " way.

As a matter of fact, the only recognised standard had been made in 1957 : it was the NF Z 44-070 : "Catalogue alphabétique de matières", which gave some general rules about the choice and the presentation of subject headings, from the card catalogue point of view (for ex. current rather than learned form for scientific words; use of the singular, and so on). As this standard only gave principles and not any choice of terms, it had frequently been associated with the "liste des vedettes-matières de : Biblio", made by the bibliographic service of Hachette (1st ed. in 1954). But it was often the card catalogue itself with the existing headings, which was the authority, and the coherence was not always ensured. There was a risk to spread the references under too numerous headings.

The adoption by great libraries (the Bibliothèque publique d'information since 1974; the Bibliothèque Nationale in 1980) of Laval University's "Répertoire de vedettes-matières", and the publication by these institutions of authorithy lists (printed lists for the BPI, lists on microfiche for the BN) have modified the scenery of subject indexing in France. In fact they have raised the problem of the participation of french libraries in a common system ; this problem is particularly important as it is connected to the question of automation and networks.

We expect that, in the projects of exchange of bibliographic data, the subject headings should have the same structure, and that the same indexation language will be used.

The number of various libraries using the BPI-BN authority lists has increased, but all are of a rather important size : scientific libraries (Bibliothèque del'Ecole polytechnique since 1979, Bibliothèque de l'Ecole Centrale), great public libraries (as Lyon and Grenoble), academic libraries (as Montpellier and Orsay).

We shall see first how the Association française de normalisation has taken part in this evolution, then which are the tools for subject indexing proposed to french libraries, and finally we shall see the question in the broader contert of the on line bibliographic data bases in project.

I - THE STANDARDIZATION of SUBJECT INDEXING : a double direction

The evolution described has led to the constitution, in april 1983, of an AFNOR committee. It regrouped members from the two directions of the ministries responsible for french libraries (Direction du Livre et de la Lecture at the Ministère de la Culture, Direction des Bibliothèques, des Musées et de l'Information scientifique et technique at the Ministère de l'Education nationale), from the Cercle de la Librairie, from the Bibliothèque Nationale, the Bibliothèque publique d'information, other libraries, from the Ecole Nationale Supérieure de Bibliothécaires, and so on.

It had assigned it-self at the beginning a double purpose : - to revise the 1957 standard " Catalogue alphabétique de matières ", too old and lacking in precision ; - to present the structure and the directions for use of the subject authority list. The BPI and the BN, which had sometimes got into different habits, because the BPI had adopted the Laval University's Repertoire sooner, agreed on a common presentation of the authority list in a " fascicule de documentation " written at the end of 1983.

In january 1984, the committee realised that it was not possible to combine in the same document the revision of the 1957 standard and the presentation of the authority list : the vocabulary and the syntax of the list were not compatible with the procedures of the standard. So it decided to let coexist two standards : one about the subject authority list; the other about the analytical subject catalogue according to the 1957 revised standard.

1 - The "fascicule de documentation " refered to has resulted in a project of standard (NF Z 47-200 : Liste d'autorité de matières : structure et mode d'emploi) , submitted to a public enquiry in France from february to june 1984 : this means that it has been sent to different institutions for comments.

Some belgian colleagues, from the Centre de Lecture publique de la communauté française de Belgique, have answered to the enquiry. The standard has been reviewed according to the answers and ratified ; it is now published and distributed. We can quote its introduction : " La présente norme recommande, à l'intention surtout des bibliothèques possédant un fonds encyclopédique et/ou voulant se référer à une liste d'autorité, l'usage d'un système reposant sur une structure et une syntaxe, et recourant à une liste de mots-matières : la liste d'autorité

matière de la Bibliothèque de l'Université LAVAL à Québec, elle-même héritée de la liste de la Bibliothèque du Congrès... Dans les bibliothèques qui l'adopteront, la cohérence du vocabulaire d'indexation sera facilitée et garantie, et l'on évitera la dispersion des notices sous des rubriques trop nombreuses. Au niveau national, et dans la perspective de l'informatisation des bibliothèques, cette liste permettra à ses participants d'accéder par un langage documentaire commun aux informations sur des fonds très étendus et divers."

The standard will be completed by a manual for subject indexing, as a sort of practical guide.

2 - The working group which has prepared the revision of the 1957 standard is inspired by a rather classical philosophy of subject indexing ; it wants to attend the small and medium sized libraries, which are afraid by a total transformation of their existing catalogues. And it is true that, if you adopt an authority list, you are nearly obliged to close one catalogue and open another, which is really terrific for librarians.

The revision of the standard is justified in the foreword by the introduction of automation and the need for more precise rules in subject indexing.

" La démarche est strictement analytique. Elle se traduit par le classement alphabétique des entrées, dans une liste non hiérarchisée. La création de relations entre termes... doit être maîtrisée afin de ne pas transformer la démarche analytique en approche systématique."

The standard gives again the classical rules (for ex. the use of the singular), and details the structure of subject headings, with different possible subheadings. It precises too the way of indicating the relations between two headings and of establishing references. This revised standard is to be submitted to a public enquiry, under the title " Indexation alphabétique par matière ". It keeps for the moment its former number 44-070, which links it to the cataloguing standards. On the contrary, the standard about the subject authority list is linked by its number (47-200) to the thesauri standards. The revised standard recommends to mind with the coherence of selected headings and then use a list giving a controlled vocabulary.

II - BIG and SMALL BIBRARIES : Two subject indexing methods?

1 - The Cercle de la Librairie has published in may 1984 a book entitled : " Choix de vedettes-matières à l'intention des bibliothèques " by M. BLANC-MONTMAITEUR et F. DAGNET. The term of "choice" is clear : it is " une liste encyclopédique de dimensions restreintes ", " une liste de mots sélectionnés pouvant servir de liste de référence aux bibliothèques publiques ". It comprises 12.500 entrées, nearly all common nouns.

The origin of this book has been the creation of a subject catalogue for a new library ; as the authors explain it in their introduction, they have chosen the construction " a priori " of a list of terms, with the help of dictionaries ; they have then compared this list with an old catalogue reflecting important holdings, at the Bibliothèque universitaire de Nanterre. This book wants to offer " à la fois un ensemble de vedettes-matières de référence et une méthode d'indexation ", method which looks like being compatible with the principles of the revised standard.

One important element is the fact that the Cercle de la Librairie, publisher of this book, is going to use it to add a subject index to its annual cumulative volume : " Répertoire Livres-Hebdo : un an de nouveautés ". For the moment this Répertoire has only title and author indexes. The current french commercial bibliography would then comprise, since the 1985 cumulative volume, a subject index with a vocabulary coming from the " Choix de Vedettes-matières ".

2 - The other indexing system, the authority list, is distributed under different forms :

- The authority list itself exists in printed volumes, different for common nouns and proper nouns for the BPI, and in microfiche for the BN, with common nouns and proper nouns in one alphabetical order, kept up-to-date by cumulative lists issued quaterly.

- The subject cataloguing made with the help of the list appears in the official bibliography : since may 1984, the Bibliographie de la France comprises subject headings under each record and recapitulates them in a subject index. There is also a subject index in the cumulative indexes.

The increasing number of its users and the need to maintain its evolutionary character raise the question of a common management of the authority list : it is necessary to avoid the risk of a too indi-

vidualistic use of the list by each library.

So the Direction des Bibliothèques has created since october 1983 a "Cellule nationale de coordination de l'indexation-matières". It manages the "liste d'autorité de matières encyclopédique, collective et hiérarchisée" : LAMECH. The members of this group have taken part in the meetings between indexers of the BN and the BPI, who intend to write the manual for subject indexing and unify the lists of subdivisions. They are also in charge of the relations with the foreign institutions which produce ou use authority lists.

To fulfill the double purpose of the list, that is to be evolutionary and collective, the Direction des Bibliothèques has created an online database; the matter is to keep the coherence of the list with many different users and to make as quickly as possible the creations and the modifications of terms.

" Le noyau de base de LAMECH est constitué par la liste d'autorité matière de la Bibliothèque Nationale afin d'assurer la cohérence avec les vedettes-matières figurant dans les notices de la Bibliographie de la France ". Searches and creations are made on line.

The create headings - common nouns, the system, with the help of electronic mail, will be as follows : the users will propose - if they have not found any heading in the more recent lists of Laval or Library of Congress - new headings, with their sources. These creations will have to be ratified by the Cellule nationale de coordination, helped by experts.

The database is on line in a national host (the SUNIST : Serveur universitaire national pour l'information scientifique et technique, near Lyon) , either by a teletype or by a videotex terminal, with a software intended for the management of dictionaries, ALEXIS ; it allows searches with boolean and proximity operators.

LAMECH as an aid to indexing database will allow a quick updating for users' network and the edition of products according different criteria.

III - ON LINE ACCESS TO BIBLIOGRAPHIC INFORMATION

We have just now talked about the changes in the scenery of subject indexing in France. They are linked to a more global phenomenon, which is the national bibliographic database. This one involves many partners : the two ministerial directions : the DBMIST - Direction des Bibliothèques, des Musées et de l'Information scientifique et technique) with the network of academic libraries ; the Direction du Livre, with the network of public libraries ; and two important producers of records :

the Cercle dela Librairie for the current commercial bibliography and the Bibliothèque Nationale for the copyright deposit and the national bibliography. It is very important for the unity of bibliographic information in France that these partners agree, but it is not always easy.

I - The project of the Cercle de la Librairie is to create a database with, at the beginning, only the new books (since october 1984), and afterwards the whole of books in print. The bibliographic service of the Cercle is cataloguing now according to the ISBD and a compatible MARC format. It creates medium records consistent with an AFNOR recent standard for descriptive cataloguing : Z 44-073, "Catalogue des monographies. Rédaction d'une notice moyenne de description bibliographique". There is also a new standard, Z 44-072, describing the minimum record.

For subject indexing, it is the vocabulary of the "Choix de vedettes-matières" which is used.

The Cercle database has not only a bibliographic aspect with the advantage of a very quick information : it has also a commercial side : booksellers will be able to make acquisitions on line. Experiences with the data service on pilot-sites could begin at the end of 1985.

2 - The Bibliothèque Nationale is implementing in its services the GEAC system to manage its bibliographic database. This one contains the records of printed documents received by the BN through copyright deposit (the Bibliographie de la France is automated since 1975), through acquisitions, donations or exchanges. It will have, at the end of 1985, 250.000 records for french monographs, 50.000 records for serials, 30.000 records for foreign books, with an important number of authority records. The goal for 1987 is 1 million bibliographic records and 500.000 authority records.

3 - The Direction du Livre has created its own software : LIBRA (Logiciel intégré pour les bibliothèques en réseau informatisé). It has recently given to the "Centre national de coopération des bibliothèques publiques" created at spring 1985 in Massy near Paris, the responsibility for the management of the database created with this software. This database is made by the network of the regional public lending libraries and of some other public libraries; it could be used too by technical regional cooperating services which would choose Libra. It gives also the localisation of documents and so is an union catalogue.

It is to hope that the producers of records and the managers of networks agree to allow users to get the best access to bibliographic information. The characteristics of the systems are different : a commercial aspect for the Cercle database; a retrospective and full cataloguing for the BN database, shared cataloguing for the LIBRA database.

In the light of this general background, we have to come back to our problem of subject indexing, and to the existence of two different indexing systems.

The Cercle de la Librairie will give in its database subject headings coming from the ' Choix de vedettes-matières".

On the other side, the Direction des Bibliothèques recommends the use of the authority list to the academic libraries which are automated. The BN and the BPI will of course use the authority list. The LIBRA database could offer the two types of indexing.

The coexistence in France of two different systems can seem unfortunate. We can also see in this fact a choice offered to libraries according to their size and the importance of their holdings.

The availability of the authority list as a database (LAMECH) will probably allow a better use of the list by the libraries participating in the network, and especially quick and unified creations of headings.

However that may be, after a relative stagnation till the end of the seventies, subject indexing recovers some vitality, due especially to the need for subject access in on line systems.

BIBLIOGRAPHIE

- ASSOCIATION FRANCAISE DE NORMALISATION - Paris - Catalogue alphabétique de matières : norme NF Z 44.070 - Paris : AFNOR, 1957
(Norme en cours de révision)

- ASSOCIATION FRANCAISE DE NORMALISATION - Paris - Liste d'autorité de matières - Structure et règles d'emploi : norme NF Z 47-200 Paris : AFNOR, 1985

- ASSOCIATION FRANCAISE DE NORMALISATION - Paris - Catalogage des monographies ; description bibliographique moyenne : fascicule de documentation Z 44 073 - Paris : AFNOR 1984

- BIBLIOTHEQUE de l UNIVERSITE LAVAL, QUEBEC - Répertoire de vedettes-matière. - 9ème édition - Québec : Bibliothèque de l'Université LAVAL; 1983 - 4 vol.

- BIBLIOTHEQUE NATIONALE - Paris - Liste des vedettes-matière - Paris : Bibliothèque Nationale, 1982→ (microfiches trimestrielles cumulatives)

- BIBLIOTHEQUE PUBLIQUE d'INFORMATION - Paris - Autorité-matière : noms communs - 5ème édition - Paris : Centre national d'art et de culture Georges Pompidou, 1981 - 4 vol.

Autorité- matière : noms propres - 4ème édition, Paris : Centre national d'art et de culture Georges Pompidou, 1981, 2 vol.

- BLANC-MONTMAYEUR (Martine) et DINSET (Françoise) - Choix de vedettes-matières à l'intention des bibliothèques - Paris : Cercle de la Librairie, 1984 -

BEST COPY AVAILABLE

IFLA GENERAL CONFERENCE, CHICAGO

Division: Bibliographic Control

Section: Classification

1985

113-CLAS-4-E

Classification in Online Systems: Research and the North American Perspective

by

Nancy J. Williamson
Faculty of Library and Information Science
University of Toronto

KEYWORD(S):

For internal use only:

To be brought to meeting no:

153

Remarks:

to be presented orally ☒

for distribution only ☐

Abstract

An analysis of the potential for a role for classification in online catalogue. Progress and problems in the North American context. Suggested areas of research on the use of classification in the design, development and use of online systems. A first step in the development of international guidelines and standards.

Introduction

In the broadest sense, classification in some form is an essential component of all document and information retrieval systems. Knowledge itself is a product of the categorization and organization of data. Document classification systems are designed to provide systematic access to documents and their surrogates with a supplementary and complementary approach being provided through alphabetical catalogues and indexes. Moreover, classificatory structures are deliberately incorporated into indexes, thesauri and other alphabetical tools for accessing information. As an aid to retrieval, classification has always played an important role in manual information systems and it should be assumed to have at least equal significance in their online counterparts. However, the use of classification should be suited to the system and, as yet, there are many unanswered questions as to the most effective methods of introducing classification into online systems. As a major project the IFLA Standing Committee on Classification and Subject Cataloguing has accepted the challenge of endeavouring to identify, foster and encourage the research and development needed to provide answers to some of those questions. What kind of research is needed? What facilities are there for carrying it out? Are there areas where standards are needed? Can these be developed within the IFLA framework? These are but a few of the questions which will be posed over the next several years.

This paper describes some of the first steps which have been taken and which are intended to lead to further investigation into classification and its role in online systems, including

online catalogues, bibliographic databases and full text retrieval systems. Obviously such a hugh topic must be divided into manageable subtopics and each subtopic handled separately. Study could begin with any one of these three areas, but the online catalogue seems to be a logical starting place for several reasons. First of all, unlike bibliographic databases, online catalogues are still in their infancy. Nationally and internationally accepted standards have not yet been developed and an atmosphere of flexibility and openness to new ideas and methods still exists. It is important to realize that this atmosphere may be somewhat more restrained several years into the future. Secondly, computer technology holds out possibilities for the use of classification in subject catalogues in ways which hitherto have not been feasible. Moreover, there is a growing climate of interest in the future of classification online. Most importantly, for the benefit of catalogue users it is desirable to have catalogues which are as effective as possible and which achieve some degree of uniformity across institutions. Designers of online catalogues are in need of reliable research and acceptable standards upon which to base their decisions. Thus, it is essential that research and development on the role of classification in online catalogues be carried out while those pioneering in such catalogues are still in a position to be responsive to new and innovative ideas and before they become bound by decisions which could close doors to the implementation of significant research findings.

Given this situation, I have begun to study the North

American library and information scene and to consider ways in which we might contribute to research in this area. Both basic and applied research are needed, and while existing catalogues may provide the bases for some studies, experimental studies will also be of great importance.

The Study

At the time of writing, this study has gone through two phases. First of all, an exhaustive literature search has been carried out, looking at four major areas - the role of classification in retrieval, subject access in library catalogues with emphasis on online catalogues, potential uses of classification in future information systems, and needed research in classification. Secondly, questionnaires were prepared and sent to North American libraries known to already have, or to be in the process of developing, online catalogues. A second set of questionnaires is being directed toward a selected number of vendors who advertise that they will install such catalogues.

The questionnaires have been designed to collect data on the degree of use being made of classification in North American online catalogues, to determine 'how' classification is being used and to identify reasons for its inclusion or exclusion in connection with these tools. Libraries have also been asked whether they would be willing to permit their online catalogues to be used as databases for further research. Some thought has also been given to problems involved in developing experimental research projects. Major obstacles are the lack of suitable databases and the absence of the required software for proceeding with experimental research. While full analysis of the data from

the questionnaire is still in progress, some trends are evident and it is possible to project some ideas for research which I would like to share with this audience and which I hope will become part of the IFLA project. First let us look at the online catalogue as it exists in North American libraries and the reasons for their rather slow response to classification as an important tool in the improvement of subject retrieval of documents.

Classification as a Device in the Online Catalogue

Historical development clearly indicates that the potential for the role of classification in online catalogues has only recently been fully recognized by North American libraries. The earliest studies[1, 2, 3] of the problems of subject access in online catalogues, concentrated on the development of strategies which would make better use of subject headings, on the enrichment of subject access with keywords, and on the promotion of the Boolean search capabilities which have already proven successful in online databases. Many of the earliest online catalogues[4] made no provision for the use of classification in retrieval, but as time has progressed, more and more catalogues are being designed which have "call number access". By and large this is a by-product of the development of online circulation systems, but there is considerable confusion over the difference between the terms "call number" and "classification". In both cases the notation is regarded primarily as a location device not as a mechanism for systematic scanning of records or the planning of search strategies. While the potential for classification as

a search tool in online systems has long been recognized by classification theorists,[5, 6, 7] the practitioners and systems designers have only recently begun to ask the right questions. The major research study now being conducted by Forest Press and OCLC under the sponsorship of the Council on Library Resources is a giant step in the right direction.

My survey indicates that there are several reasons for the course of events thus far.. In the initial stages of development of online catalogues libraries tended to avoid the unknown for fear it would increase, unnecessarily, both the initial and on-going maintenance costs of these catalogues. There was also the fear that the addition of classification as an access point might unduly delay the final implementation of the new catalogue. But perhaps the most compelling reason for omitting classification from catalogue design decisions was the historical perception that classification is a tool for the useful arrangement of documents on shelves, rather than a device of the catalogue, and the hope that the power of Boolean search could negate any need for classification as a search mechanism. To these reasons might be added a degree of dissatisfaction with document classification and its inherent weaknesses of outdatedness and inflexibility. Indeed there are those who assume that the online catalogue may make it possible to abolish classification as a means of ordering collections. It is of the utmost importance that research be conducted which will either refute or support such assumptions.

North America's slowness to respond to the potential role of classification in online catalogues was entirely predictable and can be attributed to several factors. The coordination of the

alphabetical subject catalogue with the widespread use of open stack libraries does much to encourage the use of classification as a location device, while also discouraging its use for more sophisticated strategies for searching. Recent research findings[8] indicate that as many as two-thirds of subject catalogue searches are conducted for the sole purpose of identifying call numbers which can in turn be used to locate specific items or groups of items on library shelves. Almost no thought is given to the more subtle role of classification as an instrument in the organization of knowledge. Failure to study classification as an intellectual process or to capitalize on its strengths can be attributed to a North American emphasis on speed, economy and standardization as opposed to intellectual excellence. For example, one of the most elegant uses of classification is the classified catalogue. However few North Americans have ever seen a classified catalogue. There were three or four major catalogues of this kind on the continent in the first half of the 20th century but most have ceased to exist because they are too expensive and too time consuming to maintain and they do not make the most efficient use of centralized and shared cataloging records. Many libraries get from 80 to 100 percent of their cataloging from some central agency or shared database. This is not an asset which can be ignored. In the 1950's the National Library of Canada established a classified catalogue as the best means of providing for its bilingual needs. Inevitably, after a few years, the costs of maintaining the catalogue became prohibitive and deviation from the standard alphabetical subject head-

ing approach hindered the library's ability to participate in the international exchange of cataloguing data. For these reasons it was deemed necessary to give up a tool which was intellectually useful, but economically and politically difficult to justify.

In another kind of cost cutting measure, during the 1950's and the 1960's many North American libraries discontinued the inclusion of "see also" references in their alphabetical catalogues. Again the reasoning was that, while see references might be cheap and useful, see also references were too expensive and too labour intensive to maintain. Moreover they required the expensive upkeep of subject authority files. Unfortunately, catalogue users, for the most part, were unaware that they were being deprived of the classificatory relationships among terms and headings which results from the syndetic structure created by subject references. Nor were many users aware of the availability of the volumes of the Library of Congress Subject Headings on the top of the card catalogue or at a catalogue information desk. If catalogue users were familiar with the practice, could they use the lists themselves? Would the terminology in the current edition of LCSH correspond to the heading used in the catalogue? Those large red volumes are very complex and not fully understood even by some librarians. Pity the poor user.

New systems were also rejected for economic reasons. In the early 1970's the development of the PRECIS system of indexing at the British Library brought new hope for an effective alternative to Library of Congress Subject Headings. As a system founded on the principles of classification, PRECIS would have brought a logic and systematization to alphabetical subject access to in-

formation which had previous not existed. However, this was not to be - at least not on a national or a continental scale. While there are some specialized information systems in North America which use PRECIS, such as the one developed by the National Film Board of Canada for its Film Canadiana, these systems are few in number. To its credit the Library of Congress recognized PRECIS as a system with tremendous possibilities, but time, money and staff requirements made it impossible for LC to adopt PRECIS and foster its implementation and development in North America. However this system needs re-examination in terms of its potential for online searching. Its classificatory structure has very important potential for the online catalogue.

Will history repeat itself in the online catalogue? Thorough and reliable research of the right kind could prevent this from happening. Moreover there are some hopeful signs that some efforts are being made to avoid the mistakes of the past.[9] On the North American Library scene there is a willingness to at least consider new approaches to subject access in online systems.

The capacity for maintaining computer files in online systems is clearly recognized and being well utilized in North America. As a result, see also references have been brought back into the catalogue, and subject authority files linked to the online catalogues are providing very useful browsing capabilities in some of the very newest catalogues. Furthermore the updating and maintenance of files can be quick, easy and efficient. Also classification is beginning to appear as a component of some

9

online catalogues, although as yet its uses are quite limited and unimaginative. In most cases classification is still regarded as a location device with limited provision for browsing. On the other hand there are those who continue to feel that classification has no place in the catalogue. Critical of the obsolesence and inflexibility of book classification, they point to the power of the online databases and their success with Boolean operators and suggest that this is a simple and far more effective device than classification could ever be. But is it? Well known authorities such as Charles Hildreth[10] and Phyllis Richmond[11] think not. This, and many other problems related to information retrieval, needs more study and research. Each system has its own advantages and disadvantages, but users needs are different and they have very diverse ways of tackling information problems. Therefore it is probable that most information systems need to provide users with alternative ways of gaining access to information. Surely classification is one of those ways.

Directions for Research

We can assume, I think, that classification will survive in information systems, including online catalogues, for some time to come. The question which must be answered is how will it survive? What is our best course of action? Some action is already beginning to take place.[12, 13]

In setting up a scenario for research into the role of classification in online catalogues we need to consider at least 3 points of reference from which research findings could be developed. These points of reference are the systems designer, including the classificationist, the librarians and subject cata-

loguers and the information seekers. Some of this research may be applied; some will undoubtedly be experimental. The standard classification systems, the Dewey Decimal Classification and the Library of Congress Classification must be part of this, but research must reach beyond this to examine new classification systems and new kinds of hierarchical and polyhierarchical structures which might supercede the present systems.

From the point of view of those designing systems investigation is needed into the use of computers in developing and maintaining systems online. For the standard schemes the best research potential surely exists with the publishers themselves. I am assuming that the present study being carried out by Forest Press and OCLC will do this. There will be similar needs for the Library of Congress classification and this should be done by the Library of Congress, or by a research team with LC's cooperation and blessing. Experimentally there are other lines of research which might be pursued. One important aspect is the exploration of the effective use of thesauri in online systems. What is the best kind of thesaurus to develop for online systems? Do the thesaurofacet types of systems perform better than those thesauri without the hierarchical classification component? Are tree structures such as those used in Medical Subject Headings (MESH) a mechanism which should be explored further? Experimental research should also be used to examine the possibilities of multidimensional and polyhierarchical classification. It is possible that this approach could lead to systems whereby a user might be permitted to explore relationships among terms by requesting a

topic and through an explode command be able to view the hierarchical and associative relationships with other topics. This is already being done with thesauri in bibliographic databases such as "Wilsonline", why not, also for classification in the online catalogue? This kind of classificatory system might be a more effective mechanism in an online catalogue than the traditional hierarchical display of DDC or LCC.

Moving to the needs of the classifier and the subject cataloguer, it has been suggested that we might classify online. At present, this does not appear to be a very efficient use of a computer terminal, but this could prove to be wrong. It would be a good way for classifiers to view updated schedules. Hence, research should be carried out to examine the feasibility of providing the schedules online with the ability to access them through term searching of indexes. One requirement of this would be a standard communications format analogous to the MARC formats developed for other kinds of machine readable records. Other possibilities which might be explored are the ways in which online systems could be used to update the classification in a particular library's collection. Is it possible to have a classification authority control file which is linked to bibliographical records and which would permit global search and change of classification numbers in an online catalogue? If so, perhaps this is a way of bringing the classified catalogue into online systems and keeping them up to date. If such a system were possible, would there still be need to classify books for shelf arrangement? While I do not believe that it is possible to abolish classification, it is possible to redirect it.

Ultimately all of the research must focus on what the information seeker needs and will use. For the present the standard systems used in North America seem likely to survive but we need to know more about their behaviour in online systems. As already mentioned a very thorough study of DDC in online system is in progress. A similar study should be carried out on the Library of Congress system. In the past there has been an assumption on the part of some that the DDC system is more amenable to online systems than LCC, by virtue of the hierarchical notation in DDC, but this is not necessarily so. Research which would do some comparative studies is very much needed. Should we use either of these systems as an online tool, other than as a location device? Would some kind of faceted system be better online? Another area is the whole question of display. If we were to display a classification scheme online for the user, what would we display? How would we display it? What possible role would transparent systems play in displaying classification online?

One of the great difficulties of doing applied research in North America in the area of classification is the lack of suitable databases for research and experimentation and the software needed to carry it out. We can build castles in the sky but the ultimate concern is "use". What will the information seekers use? Will they look at a classification scheme displayed on the screen and make intelligent use of it? Possibly, but possibly not. We don't know. Moreover, we should beware of dumping systems designed for manual use into computerized systems without scrutiny. It is important to develop systems which make the best

use of the technology, while serving the information-seeking public effectively.

Should the classified catalogue be revived? We don't know whether the user will use it and, in North America it is doubtful that we have one to test. The first step is to examine the use of classification in online catalogues which permit shelflist type browsing to determine how the classification is used or might be used. This could then be a starting point to find out whether these same users could be encouraged to use the potential which a classified catalogue could provide.

Beyond the online catalogue the second stage is to look into bibliographic databases. An even more exciting challenge is the question of the place of classification in non-bibliographic databases. Many have assumed that if you have the full text you don't need structured or classificatory access. In many books the index is a classification structure. Why shouldn't we need classification with full text information systems? The question is not should there be classification but how should it be incorporated. Above all it is imperative that any research be open minded, that it seek the most effective methods of using classification in online systems and that it avoid the entrenchment of existing systems which do not optimize online access to information.

Ultimately the research into the role of classification in online systems of all kinds must lead to the development of any standards needed to support classification as a tool of subject access in such systems. Compatability and flexibility in computerized information systems will continue to remain an impor-

tant factor in the international exchange of bibliographic and textual data. Standards, both national and international, have an essential part to play in this.

It is easy to ask questions but it will be more difficult to find answers. We have to start somewhere and I hope we are on the way.

References

1. Carol A. Mandel and J. Herschman, "Online Subject Access - Enhancing the Library Catalogue," Journal of Academic Librarianship 9 (July 1983): 148-155.

2. Karen Markey, Research Report on the Process of Subject Searching in the Library Catalog: Final Report of the Subject Access Research Project, Dublin, Ohio: OCLC Online Computer Library Center, Office of Research, 1983.

3. Keith W. Russell, ed., Subject Access Report of a Meeting Sponsored by the Council on Library Resources, Dublin, Ohio, June 7-9, 1982, Washington, D.C.: Council on Library Resources, 1982.

4. Stephen R. Salmon, "Characteristics of Online Public Catalogs," Library Resources & Technical Services 27 (Jan./Mar. 1983): 36-67.

5. Elaine, Svenonius, "Use of Classification in Online Retrieval," Library Resources & Technical Services, 27 (Jan./Mar. 1983): 76-80.

6. Phyllis A. Richmond, "Futuristic Aspects of Subject Access," Library Resources & Technical Services 27 (Jan./Mar. 1983): 88-93.

7. Karen Sparck Jones, "Some Thoughts on Classification for Retrieval," Journal of Documentation, 26 (June 1970): 89-101.

8. Karen Markey, Ibid.

9. Janet Swan Hill, "Online Classification Number Access: Some Practical Considerations," Journal of Academic Librarianship 1 (1984): 17-22.

10. Charles Hildreth, "To Boolean or not to Boolean," Information Technology and Libraries 2 (Sept. 1983): 235-237

11. Phyllis A. Richmond, Ibid., p. 92.

12. Karen Markey, Subject Searching in Library Catalogs: Before and After the Introduction of Online Catalogs, Dublin, Ohio: OCLC Online Computer Library Center, 1984.

13. Karen Markey, "Subject Searching Experiences and Needs of Online Catalog Users: Implications for Library Classification." Library Resources & Technical Services (Jan./Mar. 1985): 34-51.